

# Psychological Bulletin

HARRY HELSON, Editor  
*University of Texas*

---

## CONTENTS

Visual Sensitivity to Differences in Velocity.....	ROBERT H. BROWN	89
Self-Acceptance and Self-Evaluative Behavior: A Critique of Methodology.....	DOUGLAS P. CROWNE AND MARK W. STEPHENS	106
The Construction of Unidimensional Tests.....	JAMES LUMSDEN	127
Contributions and Current Status of Q Methodology.....	J. R. WITTENBORN	132
Psychotherapy as a Learning Process.....	ALBERT BANDURA	143
Theory of Situational, Instrument, Second Order, and Refraction Factors in Personality Structure Research.....	RAYMOND B. CATTELL	160
Comments on Cattell's Paper on "Perturbations" in Personality Structure Research.....	WESLEY C. BECKER	171
Cattell Replies to Becker's "Comments".....	RAYMOND B. CATTELL	176

---

Published Bimonthly by the  
American Psychological Association

VOL. 58, No. 2

MARCH 1961

### Consulting Editors

W. BEVAN, JR.  
*Kansas State University*

R. B. BAKER  
*University of Texas*

W. R. GATNER  
*Julius Hopkins University*

J. P. GILFORD  
*University of Southern California*

W. H. HOLTZMAN  
*University of Texas*

L. J. POSTMAN  
*University of California, Berkeley*

J. B. ROTTER  
*Ohio State University*

S. B. SELLS  
*Texas Christian University*

W. A. WILSON, JR.  
*Bryn Mawr College*

The *Psychological Bulletin* contains evaluative reviews of research literature and reviews of research methodology and instrumentation in psychology. This JOURNAL does not publish reports of original research or original theoretical articles.

Manuscripts should be sent to the Editor, Harry Nelson, Department of Psychology, University of Texas, Austin 12, Texas.

*Preparation of articles for publication.* Authors are strongly advised to follow the general directions given in the *Publication Manual of the American Psychological Association* (1957 Revision). Special attention should be given to the section on the preparation of the references (pp. 50-60), since this is a particular source of difficulty in long reviews of research literature. *All copy must be double spaced, including the references.* All manuscripts should be submitted in duplicate, one of which should be an original typed copy; author's name should appear only on title page. Original and mimeographed copies are not acceptable and will not be considered. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail and to check carefully the typing of the final copy.

*Reprints.* Fifty free reprints are given to contributors of articles and notes.

ARTHUR C. HOFFMAN  
*Managing Editor*

HELEN ORR  
*Promotion Manager*

*Communications*—including subscriptions, orders of back issues, and changes of address—should be addressed to the American Psychological Association, 1333 Sixteenth Street, N.W., Washington 6, D.C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

*Annual subscription:* \$10.00 (Foreign \$10.50). *Single copies,* \$2.00.

PUBLISHED BIMONTHLY BY:  
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Menasha, Wisconsin  
and 1333 Sixteenth Street, N.W., Washington 6, D.C.

Second-class postage paid at Washington, D.C., and at additional mailing offices. Printed in U.S.A.

Copyright The American Psychological Association, Inc., 1961.

# Psychological Bulletin

## VISUAL SENSITIVITY TO DIFFERENCES IN VELOCITY

ROBERT H. BROWN<sup>1</sup>

*United States Naval Research Laboratory*

In various reviews of the literature, psychologists have stressed the dependence of the perception of motion upon a multitude of factors. Kennedy (1936), for example, indicated in his review that this dependence necessitates rigorous control in the experimental method used for measuring thresholds. The need for careful analysis and experimentation, also stressed earlier by Neff (1936), has been restated more recently by Graham (1951) and Gibson (1958). Despite the caution suggested by these reviews, analysis of data available in the literature for a specific threshold proves fruitful for application to a more general form of behavior. The purpose of the present paper is to discuss this analysis.

Visual sensitivity to differences in velocity is commonly measured by presenting two objects which move at slightly different, but constant, speeds. The least detectable difference in speed is the differential threshold for the magnitude of velocity. As an initial step in the paper, consideration of angular speed indicates that it is the basic unit of meas-

urement involved in studies of the differential threshold. Plotting differential thresholds for angular speed yields a meaningful relation to a primary variable, the speed of object motion. From these thresholds, the sensitivity is readily calculated and expressed in terms of the ratio of the threshold to the speed. As a final step in the paper, this Weber ratio for velocity is applied to tracking and other predictive behavior.

### DIFFERENTIAL SPEED THRESHOLDS

#### *Angular Speed*

Graham (1951) has described the concept of visual angle and the utility of specifying stimulus extents in terms of the angle they subtend at the eye. Similarly, the visual angle per unit time or angular speed is a basic variable in experiments concerned with the visual perception of movement. Its use facilitates the comparison of data obtained under different conditions. For example, threshold measurements made in independent experiments at varying observational distances are expressed in terms of a common measure, angular speed. In addition, the use of angular speed as a stimulus specification may be necessary for good experimental design.

In Figure 1, the axis of rotation at  $Q$  may be specified in terms of a convenient reference point such as

<sup>1</sup> This review of research on the visual perception of movement and its application to tracking and other predictive behavior has been improved by the suggestions of colleagues, especially Joseph Dougherty, Robert E. Gardner, Howard Gordon, Jr., and Franklin V. Taylor of the United States Naval Research Laboratory.

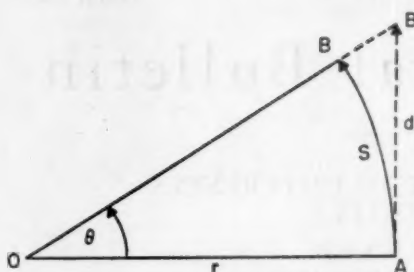


FIG. 1. Diagram representing components in the measurement of angular speed.

the front surface of the cornea. The radius of rotation ( $r$ ) is given by the distance from the reference point to the appropriate moving object. When the eye looks steadily at fixation point  $A$ , the line of regard  $OA$  is stationary. Alternatively, one may assume a rotating line of regard in experiments involving fixation on a moving object. Presently available data do not indicate unequivocally that the alternative assumptions yield a measurable difference in the perception of velocity. Fleischl (1882) reported that an object seen while fixating a stationary point moves subjectively faster than when followed by the eyes. Since Aubert (1886) confirmed the phenomenon, it has been called the Aubert-Fleischl paradox. However, the need for re-examination of the paradox is indicated by the recent work of Gibson, Smith, Steinschneider, and Johnson (1957). When they measured the accuracy of visual perception of motion, they found no difference for the two modes of observation.

As a stimulus rotates about the reference point in Figure 1, its instantaneous angular speed ( $\omega$ ) is given by:

$$\omega = \frac{d\theta}{dt} \quad [1]$$

where  $\theta$  is the angle swept by the radius vector  $r$  in time  $t$ . The value of  $\theta$  is given by:

$$\theta = \frac{s}{r} \quad (\text{in radians}) \quad [2]$$

or

$$\theta = \frac{57.3s}{r} \quad (\text{in degrees}) \quad [3]$$

The measure *angular speed* may be used advantageously not only for rotational motion but also for tangential motion. In Figure 1, the rectilinear distance  $d$  is a close approximation to the arc  $s$  for angular displacements of the magnitude usually used. For example,  $d$  exceeds  $s$  by only 1% for a  $\theta$  of  $10^\circ$ . Conversely, angular displacements less than  $10^\circ$  may be calculated with less than 1% error by substituting  $d$  for  $s$  in Equation 3. For greater displacements,  $\theta$  is calculated from:

$$\theta = \arctan \frac{d}{r} \quad [4]$$

For uniform angular motion when  $\omega$  is constant:

$$\omega = \frac{\theta}{t} \quad [5]$$

Although this equation is a special case of the earlier definition of instantaneous angular speed in derivative form, it applies with very few exceptions to experiments which have been conducted on the perception of movement. By substitution for  $\theta$  from Equation 2, uniform angular speed may be described by:

$$\omega = \frac{s}{rt} \quad (\text{in radians per unit time}) \quad [6]$$

where the arc  $s$  and the radius  $r$  are expressed in the same units. As an approximation for small angular dis-



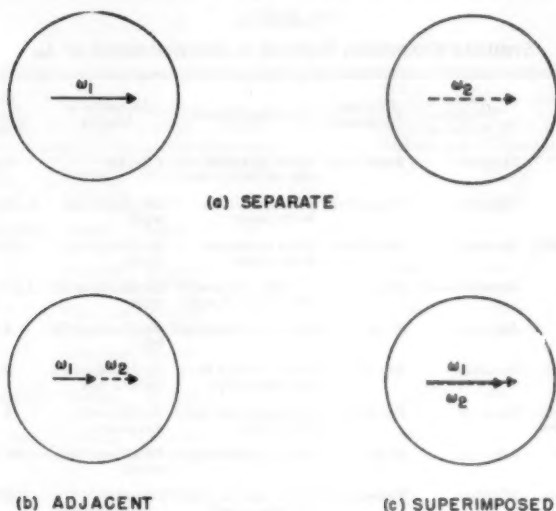


FIG. 2. Procedures used in presentations of stimulus motion.

placements, we may substitute  $d$  for  $s$  to obtain:

$$\omega = \frac{d}{rt} = \frac{v}{r} \quad (7)$$

(in radians per unit time)

or

$$\omega = \frac{57.3v}{r} \quad (8)$$

(in degrees per unit time)

where the uniform linear speed  $v$  and the observational distance  $r$  are expressed in consistent units.

#### *The Differential Speed Threshold and Its Measurement*

The differential threshold for angular speed,  $\Delta\omega$ , may be defined in terms of:

$$\Delta\omega = \omega_2 - \omega_1 \quad (9)$$

where  $\omega_2$  is a uniform angular speed an observer discriminates according to a specified criterion from the constant rate of motion  $\omega_1$ . In measurements of  $\Delta\omega$ , the spatial relationship

of  $\omega_1$  and  $\omega_2$  is critical. Three procedures used to date involve stimuli which are separate, adjacent, and superimposed. In Figure 2, a circle represents schematically an outline of a display, such as moving belt, rotating disc, or cathode-ray tube, used in presenting  $\omega_1$  and  $\omega_2$ . The speeds are represented by the vectors in each display. In Procedure *a*, the stimuli for the two speeds are spatially apart and are viewed by looking from one display to the other. In Procedure *b*, the stimuli are in immediate proximity. In Procedure *c*, they are superimposed on each other. Table 1 summarizes the most significant stimulus conditions present in measurements of  $\Delta\omega$ .

At least six experiments have been reported for measurements involving separate stimuli. Bourdon (1902) utilized two rotating white discs with a black rectangle on the edge of each. The subject adjusted the speed of one in increments until it was noticeably faster than the other. Similar meas-

TABLE 1  
STIMULUS CONDITIONS PRESENT IN MEASUREMENTS OF  $\Delta\omega$

Experimenters	Spatial Relation of $\omega_1$ and $\omega_2$	Stimulus Frequency	Stimulus Objects	Direction of Motion	Field Extent (degrees)	Observational Distance (cm.)
Bourdon (1902)	Separate	Repetitive	Black rectangle on edge of 2 white discs	Circular	6.4	200
Brown (1931)	Separate	Repetitive	Black square on white paper	Rectilinear upward	2.15-4.30	200
Brown & Mize (1932)	Separate	Repetitive	Black square on white paper	Rectilinear upward	2.15-4.30	200
Zegers (1948)	Superimposed	Single	2 needles perpendicular to line of sight	Rectilinear to S's right	3.6-15.0	15.9
Hick (1950)	Adjacent	Single	Spot on oscilloscope	Rectilinear to S's left	4.8	53.3
Ekman & Dahlbäck (1956)	Separate	Repetitive	Black vertical lines on white paper	Rectilinear to S's right or left	5.72	50
Gibson, Smith, Steinschneider, & Johnson (1957)	Separate	Repetitive	Wallpaper with pattern of dots	Rectilinear downward	8.4	122
Notterman & Page (1957)	Adjacent	Single	Spot on oscilloscope	Rectilinear horizontal	10.0	25.4
Brandalise & Gottsdanker (1959)	Separate	Repetitive	White dot on edge of 2 black discs	Circular	5.2	200

urements were made by Brown (1931) and by Brown and Mize (1932) for a black square moving upward on white paper which the observer saw in either of two windows. Ekman and Dahlbäck (1956) and Gibson et al. (1957) have made measurements involving the adjustment of  $\omega_2$  for apparent equality with  $\omega_1$ . The former utilized two apertures in each of which alternately the observer saw the horizontal motion of black vertical lines on white paper. The latter presented behind two windows a downward moving wallpaper with a pattern of dots. Most recently, Brandalise and Gottsdanker (1959) have had subjects adjust the speed of rotation of a black disc with a white dot on its edge to apparent equality with that of another. In these six experiments, the measurements of  $\Delta\omega$  were based on comparisons of the two speeds which were viewed separately in different places. Since the equipment involved rotating drums or discs, stimulation was repetitive.

Use of a moving spot on an oscilloscope has facilitated presentation of adjacent stimuli. During rectilinear motion of a pip at constant speed, an incremental change in speed is introduced. Hick (1950) and Notterman and Page (1957) measured the differential threshold in speed for a pip as it was horizontally deflected across the face of a cathode-ray tube. Temporal features of this procedure differ from the first. The stimuli are presented only once and then in immediate succession.

The procedure of superimposed stimuli may be illustrated by monocular movement parallax. When two objects move at the same linear speed perpendicular to the subject's line of sight, the difference in their angular speeds provides an indication of their distances from the subject. As the objects are brought closer together, the difference in angular speeds decreases to a threshold value. Zegers (1948) has measured the differential threshold speed for two needles by

TABLE 2  
METHODOLOGY USED IN THE MEASUREMENTS OF  $\Delta\omega$

Experimenters	Psycho-physical Method	Measure of $\Delta\omega$	No. of Speeds	No. of Subjects	No. of Measurements per Speed per Subject	Total No. of Measurements	Speed (degrees per sec.)	
							Minimum	Maximum
Bourdon (1902)	Limits	Mean	3	1	20	60	0.77	5.04
Brown (1931)	Limits	Mean	2	2	10	40	1.79	3.58
Brown & Mize (1932)	Limits	Mean	6	2-3	3-6	117	1.72	4.58
Zegers (1948) 3.6° field	Constant stimuli	Standard deviation	4	2	100	800	2.67	20.1
15.0° field	Constant stimuli	Standard deviation	6	2	100	1200	2.67	36.1
Hick (1950)	Constant stimuli	Mean	7	18	—	—	0.15	10.2
Ekman & Dahlbäck (1956)	Average error	Standard deviation	5	10	4	200	2.07	4.81
Gibson, Smith, Stein-schneider, & Johnson (1957)	Average error	Standard deviation	1	24	10	240	4.80	4.80
Notterman & Page (1957)	Constant stimuli	Mean	7	10	30	2100	0.34	22.7
Brandalise & Gotts-danker (1959)	Average error	Standard deviation	5	10	50	2500	2.7	24.3

this procedure, which temporally involves the single presentation simultaneously of  $\omega_1$  and  $\omega_2$ .

The psychophysical method used has been less critical for measurements of  $\Delta\omega$  than the spatial relationship of the stimuli. Table 2 lists significant methodological characteristics for the nine experiments. Items specially worth noting are the limited range of speeds in most experiments and the small number of measurements in some studies.

#### *The Differential Threshold as a Function of Speed*

The marked effect of spatial order may be observed by inspection of Figure 3, in which  $\Delta\omega$  is plotted against  $\omega$ . The curves and their points represent the use of adjacent, separate, and superimposed stimuli. The measure of  $\Delta\omega$  is that indicated in Table 2. Since Brown (1931) and Brown and Mize (1932) made only a small number of exploratory

measurements, the points plotted for their experiments are the geometric means of values they reported for speeds 1-2, 2-3, 3-4, and 4-5° per second. The data plotted for superimposed stimuli represent the monocular movement parallax thresholds obtained by Zegers (1948) with the widest and narrowest visual fields of the four for which he made measurements. Otherwise, the points represent all values reported in the literature for  $\Delta\omega$  as listed in Table 2.

The solid lines have been drawn with unit slope and represent a constant Weber fraction ( $\Delta\omega/\omega$ ). In the case of adjacent stimuli, solution for the intercept constant by the method of least squares yields the plotted equation:

$$\log \Delta\omega = -0.859 + \log \omega \quad [10]$$

It may be observed as a rough approximation that the differential threshold increases in direct proportion to the angular speed of a stimu-

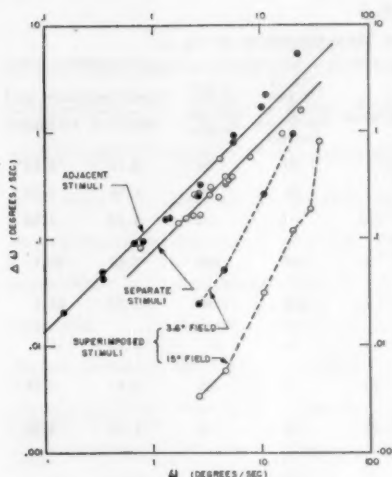


Fig. 3. The differential angular speed threshold ( $\Delta\omega$ ) as a function of the angular speed ( $\omega$ ) of stimuli which are presented spatially adjacent, separate, and superimposed.

lusion. Discrepancies in this approximation occur in a middle range of speeds (1–5° per second) where the measured  $\Delta\omega$  falls below the empirical straight line. At faster speeds,  $\Delta\omega$  increases at an increasing rate with  $\omega$ .

For separate stimuli, the least squares equation is as follows:

$$\log \Delta\omega = 1.114 + \log \omega \quad [11]$$

Under these conditions, the differential threshold increases in direct proportion to speed from approximately 1 to 10° per second. The differential threshold is greater at slow speeds, and less at fast speeds, than the best fitting straight line of unit slope.

Data obtained for superimposed stimuli can be described by a constant Weber fraction only under quite restricted conditions. Thus, for the widest field (15°) a solid line is drawn between the points for the two slowest speeds. Its equation is:

$$\log \Delta\omega = -2.893 + \log \omega \quad [12]$$

The rapid increase in the differential threshold with speed for superimposed stimuli may be interpreted in terms of instability of the retinal image and intensity effects in individual cones.

As Zegers (1948) indicates in discussion of his results, high speeds interfere with good "pickup" of the stimuli as they appear in the visual field and, also, with adequate following movements of the eyes. The influence of extent of visual field, so marked in Figure 3, was markedly decreased, if not eliminated, by providing appropriate aids during control experiments to fixation and stimulus "following." Careful measurement of the vertical distance between the curves for the 15 and 3.6° fields indicates that they could very nearly be superimposed by a shift of 0.905 log unit, the mean of separations of 0.853, 0.947, 0.909, and 0.909 log unit. We may infer that the vertical position of the curves depends primarily on stability of the retinal image. When stimulus conditions for good fixation of the stimulus are absent, the differential threshold function of Figure 3 is shifted uniformly upward with decrease in extent of the visual field.

The shape of the curves for superimposed stimuli appears to be dependent upon the intensity effects occurring in individual cones. Evidence for this inference is less direct than Zegers' control experiments involving improved conditions for fixation and pursuit of the stimulus. However, it should be pointed out that Graham, Baker, Hecht, and Lloyd (1948) measured the differential threshold as a function of the luminance of the stimulus field. Neutral tint filters were placed behind the

metal tube through which the observer saw two needles, one above the other, moving at constant and equal speeds back and forth across an illuminated field. Measurements of the precision of distance settings with one needle yielded differential angular speeds for different luminances of the visual field. The decrease in  $\Delta\omega$  as a function of the increase in luminance is described by Hecht's intensity discrimination equation upon the assumption that  $\Delta\omega$  is a measure of differences in diffraction luminances and provides a  $\Delta I$  seen against the general illumination,  $I$ .

In addition, measurements of the threshold luminance for a moving spot of light indicate that the intensity effect of speed is similar to the parallax effect of Figure 3. At moderate speeds, the threshold luminance for discrimination of motion increases in direct proportion to the speed (Brown, 1958). At faster speeds (greater than  $10^\circ$  per second), the luminance threshold increases at a disproportionate rate until it approximates an asymptote at a limiting speed (30 to  $40^\circ$  per second). This relationship, like that found by Zegers, may be interpreted in terms of intensity effects occurring in individual cones. As angular speed increases, the duration of passage of the image across a given cone is shortened. Since the intensity effect in each receptor unit is lessened, the luminance for the moving spot or the differential angular speed of the needles must be increased.

#### THE WEBER RATIO

The Weber ratio provides a convenient measure by means of which velocity discriminations may be compared with other sensory discriminations and with performance in track-

ing and predicting. The ratio of the differential threshold ( $\Delta\omega$ ) to the magnitude of the standard ( $\omega_1$ ) may be readily calculated from Equations 10-12 for adjacent, separate, and superimposed stimuli. The best estimate of  $\Delta\omega/\omega$  for an unspecified  $\omega$  is 0.138 for adjacent stimuli and 0.0769 for separate stimuli. This difference has been confirmed by Notterman (1959) in measurements made by an oscilloscope with both procedures. Since his experiment excludes variations in stimulus conditions other than the spatial order, Notterman's interpretation of the difference is of particular interest:

Subjects in the adjacent presentation case can base their discrimination on a comparison of the amount of time taken to traverse the initial and final  $1\frac{1}{2}$  inches on the scope face, or—and this is important—they can disregard time and look for the jerk which occurs when the moving spot instantaneously increases its velocity. The subjects employing the separate presentation procedure do not have this option: since the standard and comparison stimuli are separated in time, there is no jerk. In short, the subjects of the (adjacent) procedure may have changed the problem from one requiring a comparison of two velocities, to one requiring a judgment of the presence or absence of jerk (p. 3).

The marked superiority of superimposed stimuli in yielding a low Weber fraction is illustrated by the value of 0.00128 for two needles traversing an extent of  $15^\circ$  at angular speeds less than  $5^\circ$  per second. This superiority is readily understandable. Superimposition of one needle in front of the other provides an angular offset which Zegers has found to be a basic determiner of the differential angular speed threshold. The angular offset is absent when stimuli are presented adjacently in immediate succession or separately in space and time.

Variation of the Weber fraction over the whole speed range is plotted

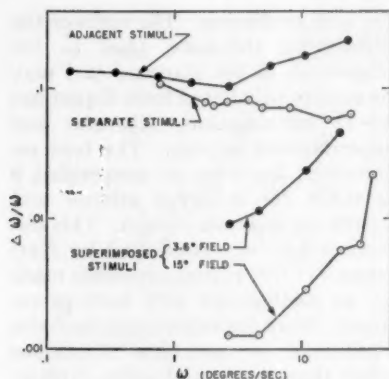


FIG. 4. The Weber ratio ( $\Delta\omega/\omega$ ) as a function of the angular speed ( $\omega$ ) for discriminations utilizing adjacent, separate, and superimposed stimuli.

in Figure 4. The points represent geometric means of values determined by different investigators at approximately the same angular speeds. Thus, the top curve for adjacent stimuli is the average of values obtained by Hick (1950) and Notterman and Page (1957). Except for the point at the slowest speed (Hick) and at the fastest speed (Notterman and Page), each point is the geometric mean of the Weber fraction in both studies. A similar procedure has been followed in averaging measurements made with separate stimuli. For superimposed stimuli, the Weber fraction has been calculated directly from Zegers' data. In this case, the ratio is directly proportional to the angular offset existing between the reference and comparison stimuli. As Zegers has indicated, the value of the angular offset (and the Weber ratio) increases with speed.

Examination of the curves of Figure 4 suggests a useful empirical generalization. The Weber fraction for nonsuperimposed stimuli is approximately constant in the mid-

range of angular speeds. Thus, in the range of 0.1 to 20° per second,  $\Delta\omega/\omega$  shows no greater change than a doubling. For adjacent stimuli, the maximal ratio is only 2.2 times greater than the minimal ratio. For separate stimuli, there is a change by a factor of 1.9. Although the Weber fraction may be fairly constant in the middle range of stimulus values, the rapid rise of the curve for superimposed stimuli suggests that the ratio may increase markedly at extremes.

The constancy of the Weber ratio for differential speed thresholds may be interpreted at a descriptive level for comparison with other sensory discriminations. Woodworth and Schlosberg (1954) have indicated that for many sensory discriminations the differential threshold is a measure of the variability of the effects of stimulation, i.e.,  $\Delta\omega = K\sigma_\omega$ . For discriminations of motion according to Brown (1960), the variability in turn is proportional to the speed, i.e.,  $\sigma_\omega = C\omega$ . It is therefore not surprising that  $\Delta\omega/\omega$  is constant, at least within limits which are not too well defined in Figure 4.

It is of interest to compare the magnitude of the ratio with that for other discriminations. Under optimal conditions, the minimal Weber fraction with superimposed stimuli is comparable to that measured for pitch discrimination with a standard tone and a comparison tone differing slightly in frequency. Measurements of pitch discrimination indicate that Weber's fraction is constant at about 0.002 beyond 250 cycles per second, rising somewhat at the lower frequencies. The differential speed threshold ratio, as measured with separate stimuli, is comparable to the Weber fraction for lifted weights. When measured by lifting weights



successively with one hand, the Weber ratio is approximately 0.075 for weights greater than 200 grams. We may conclude that the Weber ratio for differential speed thresholds not only is constant in a medium range of stimulus values, but also is of the same order of magnitude as that found for other discriminations.

#### TRACKING BEHAVIOR

Studies of tracking behavior illustrate an application of the Weber ratio for differential speed thresholds. This application is of particular interest since earlier reviews have emphasized the significant motor characteristics of tracking behavior (Birmingham & Taylor, 1954; Fitts, 1951). Perceptual characteristics have been implied by occasional observations that an operator tracks a target quickly and efficiently under optimal conditions because he estimates its present speed and acceleration and thereby anticipates its future motion. During World War II, for example, the systematic investigation of the manual controls for antiaircraft fire control systems indicated the anticipatory nature of tracking, as discussed by Helson (1949).

*Foxboro studies.* In the Foxboro studies directed by Helson, error was recorded for compensatory tracking in which the tracker tries to keep a moving pointer aligned as much as possible with a stationary reference pointer. Compensatory tracking may be contrasted with pursuit tracking in which both pointers move and the tracker aligns the following cursor under his control with the moving target pointer. In the Foxboro studies the tracker compensated for the displacement of a moving pointer, representing the aiming point, from the actual position indicated by a sta-

tionary pointer. Tracking error was measured by the time required for the target to move from its actual position to the aiming point.

Speed of the handwheel rotation was a major variable controlling tracking accuracy. For a constant-speed unidirectional course, with an increase in rate of cranking, the tracking error decreased from 55 to 6 milliseconds when a light handwheel of 2.25-inch radius was used (Foxboro Company, 1943a). Since the tracking error was consistently of the order of milliseconds and could be as small as one hundredth of the fastest reaction time, it is evident that the tracker anticipated the future motion of the target and thereby avoided the series of oscillations his long reaction time would otherwise produce.

For simple sinusoidal courses, the tracker not only anticipated the motion of the target but also used an averaging motion of the handwheel when the course was of too high a frequency to follow exactly. As course frequency increased, the tracker eliminated terminal portions of swings. Inertia in the form of a heavy handwheel or a flywheel effect smoothed the direct tracking of courses not requiring high accelerations and rapid reversals in direction (Foxboro Company, 1943b). In addition, the averaging type of behavior was dependent upon practice and familiarity with the course being tracked.

*Contemporary models for tracking behavior.* Since World War II, the concept of feedback mechanisms has been generalized to the entire field of control and communication theory in machines and animals (Wiener, 1948). As applied to antiaircraft fire control behavior, the concept states that the tracker uses the difference between the stimulus of a target's motion and

his response as a new input to make his motion correspond more closely to that of the target. Engineers analyzed human tracking performance in terms of simple servo systems with feedback (James, Nichols, & Phillips, 1947; Raggazini, 1948; Tustin, 1947). Stimulated by the mathematical systems equations which emerged from this analysis, psychologists have developed their own models to describe the behavior involved in minimizing the difference between two positions with control of one (Birmingham & Taylor, 1954; Fitts, 1951; Noble, Fitts, & Warren, 1955). These models make two basic assumptions: intermittency of response, and predictiveness of response.

*Intermittency of tracking responses.*

Despite the smooth and apparently continuous appearance of efficient tracking, experimental evidence from several sources indicates that the tracker responds intermittently. First, a time record of tracking performance shows a typical periodicity with a predominant frequency of two responses per second (Craik, 1947; Ellson, Hill, & Gray, 1947). Second, analysis of the response patterns to a step input displacement of position shows that quick corrective movements occur without visual or kinaesthetic guidance and that the typical time for completing a corrective movement, including reaction time, is approximately 0.5 second (Chernikoff & Taylor, 1952; Searle & Taylor, 1948; Taylor & Birmingham, 1948). Third, the assumption that the tracker responds intermittently at 0.5-second intervals during continuous tracking agrees with the optimal time constant obtained for conventional aided tracking (Birmingham & Taylor, 1954; Mechler, Russell, & Preston, 1949). Fourth, with the as-

sumption of 0.5-second intermittency of corrections, one may predict the optimal time constants for more complex aided-tracking control systems involving an acceleration component as well as the conventional position and rate controls (Searle, 1951).

*Predictiveness of tracking responses.*

The assumption of predictiveness in tracking responses is supported by the following findings. First, the Foxboro studies showed that the time error for manual handwheel tracking is much less than the reaction time, as discussed above. Second, pursuit tracking usually yields lower error scores than compensatory tracking (Chernikoff, Birmingham, & Taylor, 1955; Poulton, 1952; Senders & Cruzen, 1952). In the pursuit mode of tracking, responses may be made on the basis of a predictable course of the target since its marker moves independently of the marker with which the tracker follows. In the compensatory mode, prediction must be limited to the tracking error since the tracker attempts to stabilize a moving marker representing the difference between target motion and his own control motion. Third, Chernikoff et al. (1955) found that an aided-tracking control impairs performance for the pursuit mode but materially improves it for the compensatory situation. They resolved this apparently paradoxical finding by considering the nature of aided-tracking controls in terms of the predictiveness of tracking responses. With a position control, the position of the moving marker controlled by the tracker is directly proportional to the position of his control. With aided tracking, a movement of the control not only causes a proportional change in the position of the marker, but also introduces a change in its rate of motion. The aided-

tracking time constant is yielded by the ratio of the control sensitivities. With the proper time constant in compensatory tracking, the operator can correct an error with a control motion proportional to the position component of the error. He thereby sets in changes in rate of motion in amounts that are correct on the average to match the target motion. Use of the aided control in pursuit tracking requires that the tracker ignore target velocity and not attempt to predict future position. Later experiments by Chernikoff and Taylor (1957) have indicated an effect of target speed on the optimal time constant for both pursuit and compensatory tracking.

*Tracking error.* With verification of the assumptions of intermittency and predictiveness for tracking performance, it is evident how differential speed thresholds limit the tracker's responses with a position control. It may be assumed that at a given instant in time the tracker is exactly on target but that his cursor and the target are moving at different speeds. During a short period of time, the position error generated is approximately the product of this speed difference and the temporal interval. Since response intermittency holds the temporal interval constant, the tracking error is directly proportional to the speed difference which the tracker can discriminate.

Speed of target motion seems to have the same effect on tracking error as it has on the differential speed threshold as measured with nonsuperimposed stimuli, i.e., tracking error increases as a linear function of speed. Bowen and Chernikoff (1958) have investigated the relationship between magnification, speed of target motion, and tracking error with

a compensatory position-control system. Both with and without magnification, measures of tracking performance did not vary for a constant target speed when the frequency and amplitude of motion were varied over a range useful in tracking research. Tracking error increased with an increase in average speed of target motion. Departures from a linear relationship were not large.

#### PREDICTIVE BEHAVIOR

##### *Prediction Motion*

Data from Gottsdanker's series of studies of prediction motion demonstrate a marked similarity of pursuit tracking error to the differential speed threshold for adjacent stimuli. Similar to the differential threshold (approximately 14% of the speed) is the average error a tracker makes in following a target which moves at a constant speed but suddenly disappears. During the second following the disappearance, the tracker maintains the speed with an average error of 13, 14, and 16%, as measured in three separate studies by Gottsdanker (1952a, 1952b, 1955).

On some trials when the target was accelerating or decelerating at the moment of disappearance, the tracker did not continue the uniform change in speed. It should be noted, however, that at the moment of disappearance the change in speed for a 0.5-second interval was only 5 to 7% of the speed and presumably was below the tracker's threshold. Gottsdanker (1956) has reviewed the experimental literature on responses to acceleration of target motion. He concluded that smoothly accelerated motion is generally responded to as if the speed were constant, i.e., the change in speed did not exceed the differential speed threshold in the studies cited.

Gottsdanker (1952a) has measured the tracking error not only for disappearing targets, but also for completed courses. The measured error is consistent with one calculated upon the basis of the assumptions of a 0.5-second intermittency in response and a 14% speed threshold. The average error in tracking a target moving at a constant speed of 8 millimeters per second was 0.50 millimeters. If the tracker were exactly on target at a given instant, his error a half-second later would be calculated from the assumptions as the product of  $0.14 \times 8 \times 0.50$  or 0.56 millimeters, and the average error during the interval should be 0.28 millimeter. It may be assumed more realistically that the tracker was not exactly on target at the beginning of the interval. The average error should be calculated as correspondingly greater than the minimal value of 0.28 millimeter.

The prediction of tracking error from the Weber ratio for speed discriminations is not limited to visually presented stimuli, but may be extended to other stimuli. Gottsdanker (1954) has measured the precision of tapping at a constant rate of two per second. He found that subjects could maintain this rate to an accuracy of 2.4% when the stimulus of pops from a magnetic tape playback was removed. In the Foxboro studies it was found that the tracker could utilize the increased precision of rapid repetitive movements in fast handwheel cranking over the intermittent corrective responses of slower handwheel turning. As an approximation, the tracking error should be limited by the product of the repetition rate threshold and the time for each repetitive movement at the faster speeds. For example, the time error should be the product of 0.024

and 0.25 second per repetitive movement for cranking at 120 rpm. This value coincides exactly with the measured time error of 6 milliseconds for the light handwheel with short radius.

#### *Prediction of Future Positions of a Moving Target*

Although the differential speed threshold would seem to be clearly related to predictions of future position of a moving object, data on the nature of the relationship are limited. Slater-Hammel (1955) has had subjects observe a marker moving at a uniform speed over different display distances and then had them estimate when the marker would complete traversing different target distances. The display distance did not affect the error in time of estimating the arrival of the uniformly moving marker at a specified point in space. However, the error increased systematically with an increase in the target distance which the marker traversed after disappearing. In terms of percentage of the required time, the error varied between 8.9% and 21.6%. These values agree with expectations based on the Weber ratio for speed discriminations with nonsuperimposed stimuli (cf. Figure 4).

Morin, Grant, and Nystrom (1956) have reported similar results despite two important differences in their experimental procedure. First, instead of Slater-Hammel's stimulus which moved continuously at a constant speed, Morin et al. used the successive illumination of cue lights which were placed at even intervals in a horizontal row. After illumination of the last cue light, the subject estimated the time it would take the imaginary moving object to reach a

target light. Second, the object traveled at a rather slow computed speed of either 0.179 or 0.358° per second rather than the speed of approximately 5° per second used by Slater-Hammel. Results obtained by Morin et al. confirmed the fact that the error of estimating arrival increases with target distance. Significantly, they also found for their faster speed that the mean errors of estimation were generally less than 10% of the computed time. When the speed was 0.179° per second, the mean errors of estimation ranged from 25 to 53%. These values suggest an apparent extrapolation to slow speeds of the data presented in Figure 4.

Garvey, Knowles, and Newlin (1956) have measured the accuracy of prediction in terms of deviations in range and bearing between estimated and actual position plots on four different radar displays. They found that accuracy of estimated position was a function of target speed, i.e., the faster the motion of the target the less accurate the estimate. This relationship resembles that between  $\Delta\omega$  and  $\omega$  of Figure 3.

Gottsdanker and Edwards (1957) have studied a more complex type of prediction situation. Two targets moved down perpendicular paths towards an intersection but disappeared before reaching it. The subject estimated where one target would be when the other crossed the intersection. Gottsdanker concluded

that for both accelerated and constant-speed targets the prediction was based on relative positions at time of the target's disappearance rather than on relative speeds or accelerations.

#### SUMMARY

Measurements of the differential speed threshold ( $\Delta\omega$ ) have been plotted against speed ( $\omega$ ) for comparison stimuli which were presented adjacent, separate, or superimposed. As a rough approximation, the threshold increases in direct proportion to speed for nonsuperimposed stimuli over a range from 0.1 to 20° per second ( $\Delta\omega = K\omega$ ). Although the relationship for superimposed stimuli (monocular parallax) is similar, inadequate ocular following movements and receptor intensity effects modify the relationship at fast speeds (greater than 5° per second). Estimates of the Weber ratio ( $\Delta\omega/\omega$ ) of 0.138 for adjacent stimuli and of 0.0768 for separate stimuli provide a basis for interpretation of tracking and other predictive behavior. Experiments support the assumptions of intermittency and predictiveness of responses in tracking. With these assumptions, error in performance may be calculated for relatively simple tasks from the Weber ratio. For more complex tasks, constancy of the Weber ratio agrees with the linear relationship found between tracking error and speed of target motion.

#### REFERENCES

- AUBERT, H., Die Bewegungsempfindung. *Arch. ges. Physiol.*, 1886, 39, 347-370.
- BIRMINGHAM, H. P., & TAYLOR, F. V. A human engineering approach to the design of man-operated continuous control systems. *USN Res. Lab. Rep.*, 1954, No. 4333.
- BOURDON, B. *La perception visuelle de l'espace*. Paris: Reinwald, 1902.
- BOWEN, J. H., & CHERNIKOFF, R. The effects of magnification and average course velocity on compensatory tracking. *USN Res. Lab. Rep.*, 1958, No. 5186.
- BRANDALISE, B. B., & GOTTSANKER, R. M. The difference threshold of the magnitude of visual velocity. *J. exp. Psychol.*, 1959, 57, 83-88.



- BROWN, J. F. The thresholds for visual movement. *Psychol. Forsch.*, 1931, 14, 249-268.
- BROWN, J. F., & MIZE, R. H. On the effect of field structure on differential sensitivity. *Psychol. Forsch.*, 1932, 16, 355-372.
- BROWN, R. H. Influence of stimulus luminance upon the upper speed threshold for the visual discrimination of movement. *J. Opt. Soc. Amer.*, 1958, 48, 125-128.
- BROWN, R. H. Some methodological considerations in measuring visual thresholds to velocity. *Percept. mot. Skills*, 1960, 11, 111-122.
- CHEERNIKOFF, R., BIRMINGHAM, H. P., & TAYLOR, F. V. A comparison of pursuit and compensatory tracking under conditions of aiding and no aiding. *J. exp. Psychol.*, 1955, 49, 55-59.
- CHEERNIKOFF, R., & TAYLOR, F. V. Reaction time to kinesthetic stimulation resulting from sudden arm displacement. *J. exp. Psychol.*, 1952, 43, 1-8.
- CHEERNIKOFF, R., & TAYLOR, F. V. Effects of course frequency and aided time constant on pursuit and compensatory tracking. *J. exp. Psychol.*, 1957, 53, 285-292.
- CRAIK, K. J. W. Theory of the human operator in control systems: I. The operator as an engineering system. *Brit. J. Psychol.*, 1947, 38, 56-61.
- EKMAN, F., & DAHLBÄCK, B. A subjective scale of velocity. Report No. 31, 1956, Psychological Laboratory, University of Stockholm.
- ELLSON, D. G., HILL, H., & GRAY, F. Wave length and amplitude characteristics of tracking error curves. *USAF Air Material Command memo. Rep.*, 1947, No. TSEAA-694-2D.
- FITTS, P. M. Engineering psychology and equipment design. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 1287-1340.
- FLEISCHL, E. Physiologische-optische notizen. *SB Akad. Wiss. Wien*, 1882, 86, 17-25.
- FOXBORO, COMPANY. Handwheel speed and accuracy of tracking. *OSRD Rep.*, 1943, No. 3453. (Foxboro Co., PB 40615) (a)
- FOXBORO COMPANY. Inertia, friction, and diameter in handwheel tracking. *OSRD Rep.*, 1943, No. 3454. (Foxboro Co., PB 40614) (b)
- GARVEY, W. D., KNOWLES, W. B., & NEWLIN, E. P. Prediction of future position of a target track on four types of displays. *USN Res. Lab. Rep.* 1956, No. 4721.
- GIBSON, J. J. Research on the visual perception of motion and change. In *Second symposium on physiological psychology*. Washington, D. C. ONR, 1958. Pp. 165-176.
- GIBSON, J. J., SMITH, O. W., STEINSCHNEIDER, A., & JOHNSON, C. W. The relative accuracy of visual perception of motion during fixation and pursuit. *Amer. J. Psychol.*, 1957, 70, 64-68.
- GOTTSDANKER, R. M. The accuracy of prediction motion. *J. exp. Psychol.*, 1952, 43, 26-36. (a)
- GOTTSDANKER, R. M. Prediction-motion with and without vision. *Amer. J. Psychol.*, 1952, 65, 533-543. (b)
- GOTTSDANKER, R. M. The continuation of tapping sequences. *J. Psychol.*, 1954, 37, 123-132.
- GOTTSDANKER, R. M. A further study of prediction-motion. *Amer. J. Psychol.*, 1955, 68, 432-437.
- GOTTSDANKER, R. M. The ability of human operators to detect acceleration of target motion. *Psychol. Bull.*, 1956, 53, 477-487.
- GOTTSDANKER, R. M., & EDWARDS, R. V. The prediction of collision. *Amer. J. Psychol.*, 1957, 70, 110-113.
- GRAHAM, C. H. Visual perception. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 868-920.
- GRAHAM, C. H., BAKER, K. E., HECHT, M., & LLOYD, V. V. Factors influencing thresholds for monocular movement parallax. *J. exp. Psychol.*, 1948, 38, 205-223.
- HELSON, H. Design of equipment and optimal human operation. *Amer. J. Psychol.*, 1949, 62, 473-497.
- HICK, W. E. The threshold for sudden changes in the velocity of a seen object. *Quart. J. exp. Psychol.*, 1950, 2, 33-41.
- JAMES, H. M., NICHOLS, N. B., & PHILLIPS, R. S. *Theory of servomechanisms*. New York: McGraw-Hill, 1947.
- KENNEDY, J. L. The nature and physiological basis of visual movement discrimination in animals. *Psychol. Rev.*, 1936, 43, 494-521.
- MECHLER, E. A., RUSSELL, J. B., & PRESTON, M. G. The basis for the optimum aided-tracking time constant. *J. Franklin Inst.*, 1949, 248, 327-334.
- MORIN, R. E., GRANT, D. A., & NYSTROM, C. O. Temporal predictions of motion inferred from intermittently viewed light stimuli. *J. gen. Psychol.*, 1956, 55, 59-71.
- NEFF, W. S. A critical investigation of the visual apprehension of movement. *Amer. J. Psychol.*, 1936, 48, 1-42.
- NOBLE, M., FITTS, P. M., & WARREN, C. E. The frequency response of skilled subjects in a pursuit tracking task. *J. exp. Psychol.*, 1955, 49, 249-256.
- NOTTERMAN, J. M. Research on visual discrimination of velocity. Status Report



- No. 3, 1959, Princeton University, USAF Office of Scientific Research, Contract AF 49 (638)-381.
- NOTTERMAN, J. M., & PAGE, D. E. Weber's law and the difference threshold for the velocity of a seen object. *Science*, 1957, **126**, 652.
- POULTON, E. C. Perceptual anticipation in tracking with two-pointer and one-pointer displays. *Brit. J. Psychol.*, 1952, **43**, 222-229.
- RAGGAZINI, J. R. Engineering aspects of the human being as a servomechanism. Paper read at American Psychological Association, Boston, September 1948.
- SEARLE, L. V. Psychological studies of tracking behavior: IV. The intermittency hypothesis as a basis for predicting optimum aided-tracking time constants. *USN Res. Lab. Rep.*, 1951, No. 3872.
- SEARLE, L. V., & TAYLOR, F. V. Studies of tracking behavior: I. Rate and time characteristics of simple corrective movements. *J. exp. Psychol.*, 1948, **38**, 615-631.
- SENDERS, J. W., & CRUZEN, M. Tracking performance on combined compensatory and pursuit tasks. *USAF WADC tech. Rep.*, 1952, No. 52-39.
- SLATER-HAMMEL, A. T. Estimation of movement as a function of the distance of movement perception and target distance. *Percept. mot. Skills*, 1955, **5**, 201-204.
- TAYLOR, F. V., & BIRMINGHAM, H. P. Studies of tracking behavior: II. The acceleration pattern of quick manual corrective responses. *J. exp. Psychol.*, 1948, **38**, 783-795.
- TUSTIN, A. The nature of the operator's response in manual control and its implications for controller design. *J. Inst. Elec. Engr.*, 1947, **94**, 190-202.
- WIENER, N. *Cybernetics or control and communication in the animal and the machine*. New York: Wiley, 1948.
- WOODWORTH, R. S., & SCHLOSBERG, H. *Experimental psychology*. New York: Holt, 1954.
- ZEGERS, R. T. Monocular movement parallax thresholds as functions of field size, field position, and speed of stimulus movement. *J. Psychol.*, 1948, **26**, 477-498.

(Received January 21, 1960)

## SELF-ACCEPTANCE AND SELF-EVALUATIVE BEHAVIOR: A CRITIQUE OF METHODOLOGY

DOUGLAS P. CROWNE AND MARK W. STEPHENS<sup>1</sup>  
*Ohio State University*                      *Purdue University*

"Self-acceptance" has become a popular concept in psychological literature. Along with "rigidity," "authoritarianism," and "conformity," it has come to particular prominence in the last decade, perhaps reflecting an evolution in value systems in American culture. Concepts pertaining to the self have been given considerable space in the writings of personality theorists and social-personality psychologists and inevitably have found their way into psychological research.

Self-acceptance has been particularly identified with Rogers' personality theory and is accorded the status in that system of a major therapeutic goal. Phenomenological research on self-acceptance dates from the classic study of Raimy (1948). However, very similar concepts have played dominant roles in other theories—e.g., Snygg and Combs (1949), Horney (1950), and Sullivan (1953). More important, self-acceptance seems to have been pre-empted for less systematic, eclectic usage by a great many practicing clinicians and researchers (Cowen, 1956; Cowen, Heilizer, Axelrod, & Alexander, 1957; Zuckerman, Baer, & Monashkin, 1956; Zuckerman & Monashkin, 1957). The major portion of the research on self-acceptance derives from Rogers' self-theory, but

studies based on other theories (Block & Thomas, 1955; Sarbin & Rosenberg, 1955) and the generally empirical investigations referred to above attest to the breadth of current interest in the behaviors subsumed under this broadly interpreted construct.

While no single definition of self-acceptance would be accepted by all who use the term, the phenomenological view of Rogers seems to represent at least a common point of departure. From the definition of a *self-concept* construct the concept of self-acceptance is derived, referring, at least operationally, to the extent to which this self-concept is congruent with the individual's description of his "ideal self."

The majority of self-acceptance tests have followed this model (see Table 1). A somewhat different psychometric model has been proposed by Gough (1955), in which self-acceptance is inferred from the ratio of "favorable" self-descriptive statements to the total number of self-descriptive statements made by the subject.

A common denominator in the definition of self-acceptance, judging from the operations employed in its assessment, would seem to be the degree of self-satisfaction in self-evaluation. This definitional consensus, however, is achieved at the level of operations, and other meanings may be implied by self-acceptance *constructs*. Phenomenological theorists, for example, appear to be interested in an "internal" phe-

<sup>1</sup> The authors would like to express their indebtedness to the following persons, who critically read this paper and made a number of valuable suggestions: Donald Campbell, Shephard Liverant, Julian Rotter, Lee Sechrest, Charles Smock, and Janet Taylor.

TABLE 1  
CLASSIFICATION OF SOME TESTS OF SELF-ACCEPTANCE

Name of Test	Type	Score Obtained
SIO (self-ideal-other) <i>Q</i> sort (Rogers & Dymond, 1954)	<i>Q</i> sort	Pearson correlation between sorts of self and ideal on 100 items. Also, "adjustment score" based on number of favorable statements placed on "like me" end of distribution and number of unfavorable statements placed on "unlike me" end.
Index of Adjustment and Values (Bills, 1958; Bills, Vance, & McLean, 1951)	Adjective rating scale	Self-acceptance score = sum of self-concept ratings (1-5 scale) on 49 traits. Also, a self-ideal discrepancy score is calculated. Norms available.
Adjective Check-List (Gough, 1955)	Adjective check list	Self-acceptance score = number of favorable adjectives checked divided by total number of adjectives checked.
Buss scale (Buss & Gerjuoy, 1957; Zuckerman & Monashkin, 1957)	Adjective check list	Sum of differences without regard to sign of scale values (based on psychologists' ratings) of adjectives checked on self and ideal descriptions.
Self-Rating Inventory (Brownfain, 1952)	Self-rating scale	"Positive self-concept" and "negative self-concept" scores. Self-acceptance = sum of positive self-concept description weights minus negative self-concept description weights, disregarding sign.
Attitudes toward Self and Others Questionnaire (Phillips, 1951)	Self-rating scale	Sum of weights (1-5) on each item. Norms available.
Berger Self-Acceptance scale (Berger, 1952)	Self-rating scale	Sum of item weights (1-5).
Interpersonal Check List (LaForge & Suckek, 1955)	Adjective check list	Intensity scale values for each adjective (1-4). Self-acceptance = discrepancy between self and ideal ratings.

nomenal state. Other theorists (Block & Thomas, 1955) have formulated self-acceptance as a function of an ego-control construct. The phenomenological concept of Rogers and the psychoanalytic set of meanings implied by Block and Thomas' construct of ego control probably diverge in important respects. The purpose here, however, is merely to illustrate the point that emphasis on definitional clarity achieved at an operational level tends to ignore the

probably significant differences in the implied theoretical meanings of self-acceptance.

Reflecting in part the widespread interest in self-acceptance are the numerous instruments which have been devised to measure the construct. A striking phenomenon of research in this area is that these tests, characterized by a diversity of both theoretical and psychometric models, have apparently been assumed to be interchangeable. Thus,

characteristic of self-acceptance research appears to be a basic conception that measures of this construct possess face validity: that is, in a simple denotative sense, the tests are viewed as being manifestly similar (Peak, 1953).

Criterion validation of self-acceptance tests is, of course, logically impossible, and attempts at construct validation do not lend much faith in the validity even of a particular test, much less of all the different tests. Face validity, however, has apparently been assumed without question. The acceptance of face validity—that is, manifest similarity—implies adherence to a further assumption incorporated in phenomenological theory—that of the validity of self-reports (Rogers, 1951, p. 494). In terms of these assumptions, a self-acceptance test is valid if it looks like a self-acceptance test and is similar to other tests, and what a person says about himself self-evaluatively is accepted as a valid indication of how he “really” feels about himself.

The acceptance of these assumptions, whether acknowledged or implicit, has definite implications for the assessment of self-acceptance and for the interpretation of experimental results in this area. This paper will show that there are four major problems in the measurement of this construct and that, in view of the common adherence to these assumptions, the results of studies on self-acceptance are rendered highly ambiguous. These issues seem, despite their essential pertinence to research on self-acceptance, to have been sufficiently ignored to warrant exposition in this paper. It will be seen that these issues are not limited solely to self-acceptance, but represent instead basic logical and psychometric considerations which may serve to

illustrate problems in personality research in general.

#### EQUIVALENCE OF OPERATIONS

As observed above, the diverse tests of self-acceptance have been assumed to be equivalent operations for measuring behaviors subsumed under the construct. The failure of experimenters to consider the problem of the equivalence of assessment operations in published reports (Bills, Vance, & McLean, 1951; Block & Thomas, 1955; Calvin & Holtzman, 1953; Cowen, Heilizer, & Axelrod, 1955; Hillson & Worchel, 1957; Phillips, 1951) raises the question of the basis on which the findings of individual studies employing different measuring operations are generalized and incorporated in the larger body of self-acceptance research. The basis of generalization, in view of the absence of explicit consideration of the question, must be inferred to lie in the assumption of face validity as defined above. Even statements implying differences among self-acceptance tests fail to deal with the logically sequential question of the extent to which these differences may mean that self-acceptance as measured by Test 1 is not the same as self-acceptance as measured by Test 2. The following excerpt illustrates this point (Cowen et al, 1955):

Presumably each of these classes of [self-acceptance] measures has certain peculiar advantages and limitations. . . . In any case, a good many data have now been presented demonstrating some empirical validity for both types of measures (i.e., they can discriminate among subjects with respect to other personality and behavioral indices in a manner roughly consistent with predictive expectations based on phenomenological theory) (p. 242).

These writers do not make clear what relationship obtains between

the classes of self-acceptance tests (tests yielding discrepancy scores versus self-concept rating devices) or, more basically, how phenomenological personality theory can lead to operations that apparently can satisfy certain predictions in the case of one class of instruments but requires different operations to obtain positive results from other hypotheses based on the same construct.

According to the notion of face validity, what looks like a test of self-acceptance *is* such, by definition. All the test constructor is required to do, in terms of this criterion, is to elicit self-evaluative statements from subjects. All measures that conform to this requirement achieve validity and are therefore equivalent. By this procedure the test itself becomes the construct, in the sense of the narrowest kind of operational definition.

An operational definition stating what is measured by a given device or procedure in terms of specified measurement operations is, of course, a perfectly legitimate and necessary procedure in scientific investigation *as long as the interpretation of results is strictly confined to the particular test or measurement procedure*. A problem arises, however, when an attempt is made to generalize from experimental findings with a particular test to results obtained by *different* assessment operations. The problem similarly occurs in another case when a certain test is applied to an experimental problem and negative results are interpreted as disconfirming the hypotheses relating the construct to observables. As Jessor and Hammond (1957) have pointed out, in the absence of an explicit, logical relationship between the superordinate construct and the operations designed to assess it, conclusions cannot be made concerning the validity

of the hypotheses since invalid measurement operations could equally account for negative findings.

The point at issue is that tests of self-acceptance (or, for that matter, of any construct) which are based on different construct systems and in the development of which different procedures and items have been employed are not equivalent *in the absence of empirical demonstration of their relationships*; they must be shown to be either highly related to each other or similarly related to other constructs in the nomological net. Further, in the absence of demonstrated equivalence, experimental results cannot be generalized to findings with a different instrument. This seems to be so obvious a consideration that explication here is redundant. The fact remains, however, that the equivalence of self-acceptance tests has been assumed despite their independent derivation and despite the relative lack of empirical demonstration that there is a high degree of common variance among them.

In respect to the latter point, three studies are of interest. Bills (1958) reports a correlation of .24 between the self-concept score on the Index of Adjustment and Values (IAV) and the "self-score" of the Phillips Attitudes Toward Self and Others Questionnaire (1951). A correlation of .56 is reported between the Bills self-ideal discrepancy score and the Phillips self-score. Omwake (1954) found a correlation of .55 between the IAV self-acceptance (self-ideal discrepancy) score and the self-score on the Phillips questionnaire and a correlation of .49 between the self-acceptance score on the IAV and the Berger self-acceptance scale (Berger, 1952). In a recent study, Cowen (1956) found that two self-acceptance

measures yielding self-ideal discrepancy scores (Bills IAV and the Brownfain Self-Rating Inventory) were uncorrelated. The magnitude of these correlations indicates that the prediction of scores on one of these measures from scores on another would be accompanied by a wide margin of error.

The diversity of item selection procedures, item content, type of response elicited, and test format which is characteristic of test construction in this area suggests that what is operationally defined as self-acceptance on one test may be quite different from the sample of self-evaluative behavior elicited in another psychometric situation. Further, self-acceptance is construed differently by different theorists (cf. Block & Thomas, 1955; Butler & Haigh, 1954; LaForge & Suczek, 1955; Sarbin & Rosenberg, 1955), and these definitional differences are undoubtedly reflected in self-acceptance tests.

Even if one grants the assumption of face validity with its clearly implied meaning of equivalence *as made by the experimenter*, to assume that subjects will perceive these psychometric situations in the same way is another matter. It is quite conceivable that subjects may categorize the self-evaluative situations represented by the various tests of self-acceptance quite differently, with the result that scores obtained on these measures will not be congruent. According to this argument, a subject's expectancies that his goals will be achieved or frustrated as a result of his sorting a number of statements on a forced-choice distribution from "like me" to "unlike me" (Butler & Haigh, 1954) may be quite different from the expectancies aroused by a situation in which he is asked to attribute certain

adjectival characteristics to himself (Gough, 1955). Ironically, a phenomenological definition of a self-report variable is particularly obligated to account for differences in the subject's *perception* of the measurement device. In any case, unless it can be shown that there is a high degree of congruence of the various measures within the experimental populations sampled, one is without means of measuring self-acceptance as phenomenologically defined. The individual's private, unique experience of self-satisfaction or dissatisfaction remains, indeed, private.

It seems highly probable that differences among self-acceptance tests plus the likelihood that subjects will categorize these tests differently may result in the sampling of relatively nonoverlapping behaviors by the various tests. To be recognized is the fact that this is an empirical problem for which, to the writers' knowledge, the three studies cited above provide the only suggestive evidence.<sup>2</sup> The recently proposed model (Campbell & Fiske, 1959) for assessing convergent and discriminant validity would seem to be highly appropriate for determining the tenability of the assumption of equivalence of operations for measuring self-acceptance.

#### DEFINITION OF THE CONSTRUCT

##### *Specifying Parameters*

The ability to reach generalized conclusions from current self-acceptance research seems to be limited by a failure to give adequate definitions to the construct itself. As Rotter

<sup>2</sup> Since the completion of this article, further research has been published which bears directly on the problem of the equivalence of self-acceptance tests and suggests that a socially desirable response set may constitute a major source of variance (Crowne, Stephens, & Kelly, 1961).



(1954) has pointed out, it is important to distinguish between ideal, theoretical, and operational definitions of a given construct. An experimenter can define self-acceptance, for example, as the behavior sample (or as the "internal" phenomenal state *reflected* by the behavior sample) obtained on a particular test. But he is usually not interested in restricting his interpretation of his findings (if any) to this limited behavior sample, and he seeks to place his results in the larger context of research by other investigators and to generalize his findings to "real life" situations such as those encountered in clinical practice. By a narrow interpretation of operationism, the experimenter has made it logically indefensible to relate his findings to a theoretical system, to results obtained with other measurement devices, or to "real life" situations. When nothing more than an operational definition is offered, the parameters defining the variable are not specifiable, and there is no basis for generalization of the results.

At the other extreme, definitions of self-acceptance at an abstract level, not specifically articulated with other variables in a theory or tied to a specific test, are apt to be semantically loose and to be subject to differing interpretations. It is true, of course, that definitions of variables at this level transcend any particular set of operations and can usually be applied to an infinite variety of situations and behaviors. The looseness of such definitions, however, precludes rigorous tests of hypotheses and makes precise communication impossible. In self-acceptance research there have been few if any definitions of the construct which are not either rigidly operational or highly abstract.

The deduction from an abstract

definition, with all its surplus meanings, to specific operations is likely to be a tenuous one and, perhaps more often than not, is a private, nonrepeatable process. An intervening step is necessary in which the construct is broadly defined in terms of specific behavioral referents and preferably in relation to other variables in a specific theory. A "working definition," as Rotter has defined it, clearly represents an attempt to specify the parameters of the variable in question so that both generality and precise communication are gained. Self-acceptance research appears to have lacked such definitions.

Although this paper is chiefly concerned with pointing out certain methodological pitfalls in research on self-acceptance, some clarification may be achieved by defining briefly this intermediate theoretical step and attempting to relate the logic of construct validation to the more general theoretical problem. Rotter's working definition could be described as a *definition at the construct level*. In terms of this view, the behavioral referents and the *hypothesized relationships* of the construct are described as part of its definition—that is, the implied meanings of the term are publicly specified. In effect, specifying the behavioral referents and hypothesized relationships reduce to the same thing: locating the construct in a nomological net. In the language of test construction, Cronbach and Meehl (1955) write:

Construct validation takes place when an investigator believes that his instrument reflects a particular construct to which are attached certain meanings. The proposed interpretation generates specific testable hypotheses, which are a means of confirming or disconfirming the claim. . . . To validate a claim that a test measures a construct a *nomological net surrounding the concept must exist* [italics added] (pp. 290-291).

The logic of construct validation cannot be invoked to justify the identification of a particular set of operations as unique to a given construct, nor does it support the view that a construct is "validated" by the confirmation of a single hypothesis. The establishment of a single relationship belongs more properly in the domain of criterion oriented validity, as Cronbach and Meehl point out. With construct validation procedures clearly at issue, it would seem to be desirable to specify in advance the referents of self-acceptance. When the situations in which the behaviors subsumed under the construct and the behaviors themselves are identified, some idea of the generality and functional unity of self-acceptance is afforded, and relationships to other constructs, situations, and measurement operations can be suggested at a logical level.

Underwood (1957) has described the difficulty in moving from theoretical definitions (or constructs) to operational definitions—a difficulty that appears to be characteristic of psychological research. Campbell and Fiske (1959) have extended Underwood's point to show that the transition from operations to construct can involve perplexities equally difficult. The essence of the latter problem is that a single set of operations is capable of multiple interpretations; convergence on a single interpretation (that is, establishing that a relationship holds in a particular nomological net and cannot be more adequately accounted for in another net) is achieved by a process of triangulation from a number of different operations. Convergent validation, however, involves complex designs and extensive preliminary research efforts. Further, convergent validation does not necessarily help to make more explicit the descent from a theo-

retical model to measurement operations. According to the present view of definition at the construct level, this explicitness would be achieved and the reverse problem, that of interpreting results from a set of operations, might be at least partially solved. That is, alternative explanations of experimental findings could be examined in the light of the hypothesized relationships proposed in different construct systems claiming to explain the same body of data, with the result that incomplete or inconsistent interpretations might be discarded in favor of interpretations whose "fit" to the data is more adequate.

For example, phenomenological theory implicitly hypothesizes a linear relationship between self-acceptance and adjustment (Butler & Haigh, 1954), while acknowledging the possibility that very high reported self-acceptance may indicate "defensive" unwillingness to reveal personal dissatisfaction. Block and Thomas (1955), however, have shown that a curvilinear model, in which both very high and very low self-acceptance are associated with maladjustment, affords a better explanation of the phenomenon of defensiveness. It is conceivable that more explicit formulation of the phenomenological self-acceptance construct and its derived test procedures might have provided a more adequate explanation of defensive responding in the Butler and Haigh study. More precise definition of the variable in question might thus have directed a search for operations less susceptible to systematic response bias.

In a recent paper, Cowen and Tongas (1959) have reviewed a number of construct validation studies on the IAV (Bills, 1958). They point to the fact that several of these studies have reported significant results in

the direction opposite to theoretical expectation. In one study, on 10 of 21 hypotheses specifying differences between high and low self-acceptance scorers, many differences were found which indicated that subjects with high self-acceptance scores were more maladjusted than low scorers (Bills, 1953a). As Cowen and Tongas observe, high self-acceptance should theoretically be associated with satisfactory adjustment, not maladjustment. Another theoretical inconsistency occurred in the failure to show that lowered self-concept ratings and longer response times in word association are associated with conflict and emotionality. The results of this study were again, in fact, significant in the opposite direction (Bills, 1953b). Bills interpreted these findings as indicating a decrease in defensiveness. Cowen and Tongas argue, however, that:

Unless procedures can be specified before the fact, by which we can discriminate the high SC (self-concept) score representing good adjustment from the high SC score representing defensiveness, we are operating within a closed system in which the results of a given experiment, irrespective of their direction, can be interpreted as confirming the underlying theory (pp. 362-363).

Self-acceptance research is in need of clear construct-level definitions in which the relationships of the construct to other variables are explicitly stated. These definitions must refer primarily to the relationship of self-acceptance to other variables in the general theory in which the construct is embedded. Depending upon the particular theory, definitions might specify the nature of the relationship of self-acceptance to adjustment; to such personality variables as creativity, neuroticism, and defensiveness; to interpersonal variables such as acceptance of others; to environmental, social, and cultural variables

as, for example, the role of cultural sanctions in self-evaluation, or the influence of the experimental (or therapeutic) context on self-appraisal.

### *Representative Sampling*

A second problem associated with the definition of the parameters of self-acceptance concerns the representative sampling of self-acceptance test items. As applied to the construct of self-acceptance, the problem of representative sampling is involved in the systematic sampling of some specified universe of self-evaluative behaviors. Assuming that one has defined this population theoretically, it is then of importance to draw one's sample of test items in such a way as to represent their occurrence in the population. The achievement of representative sampling in this respect means that generalization can reasonably be attempted to other situations and/or behaviors than those of a particular experiment or test. Although the behavioral referents of self-acceptance might seem obvious, on closer scrutiny it appears that there is notable confusion resulting from a lack of consensus as to what these referents are.

Some examples from published research may illustrate what is implied by failure to sample representatively a population of self-evaluative behaviors. Butler and Haigh (1954) begin with Rogers' abstract definition of the self-concept. Then, they write:

A set of one hundred [self-reference] statements was taken at random from available therapeutic protocols. (Actually, the statements were selected on the basis of accidental, rather than random, sampling) (p. 57).

The population of relevant self-percepts was therefore restricted to those verbalized by some sample of

clients in client centered therapy, the basis for sampling was accidental, and thus there is no precise definition of self-acceptance in terms of what particular self-percepts define its parameters. The finding that changes in self-acceptance were demonstrated to occur as a function of client centered therapy is thereby limited to the particular conditions of this experiment, the subject population used, and the particular items employed in the *Q* sort measure. For example, it is quite possible (but unknown) that the statements used comprise a sample biased in favor of client centered counseling as perceived and defined by the judges (presumably Butler & Haigh) who selected the items.

A second example can be seen in the development of the IAV (Bills et al., 1951). The items (adjectives) in the IAV were drawn from Allport and Odbert's (1936) list of 17,953 traits. The basis of selection was the frequent appearance of the adjective in question in client centered interviews and whether it presented a "clear example of self-concept definition." Self-evaluation on the IAV, then, pertains only to the Allport and Odbert traits mentioned frequently in client centered interviews, and generalization to other self-evaluative situations, or traits, would be tenuous.

Gough's (1955) Adjective Check List (ACL) affords a third illustrative example. The ACL consists of 300 adjectives selected from Cattell's (1943, 1946) consolidation and factorization of the Allport and Odbert trait list. The basis on which the 300 adjectives in the ACL were derived from Cattell's list of 171 trait variables is not specified. In addition, the assumptions both of Allport and Odbert in their original derivation of the trait list and of Cattell in his

factorization are further restrictions in interpreting ACL scores.

With such lists of traits or items it is necessary to assume either that they truly represent *all* self-percepts, or at least that they represent the most important ones. But, especially for the phenomenologist, must it not be assumed that these are different for different subjects and/or subject populations? Must not this list, then, be tailor-made to the subject to be truly representative for him (a totally idiographic procedure)? Perhaps what is required is that the subject generate his own list of self-descriptions, or a self-description, and the values he attaches to the separate elements and to the composite. Kelly's (1955) Role-Construct Repertory Test appears to fit this model.

It would seem possible to achieve some degree of representativeness in the sampling of a defined universe of self-reference items. The definition of the population is properly referable to the theory in which the self-acceptance construct is embedded. That is, one should be able to deduce from the theory the nature of the items to be sampled (although, from a phenomenological theory, one might protest that this population of items is unique to the individual; but this only thickens the soup). Not only should the population of subjects be specifiable (for example, the theory has particular relevance to persons in client centered therapy), but what constitutes a relevant self-evaluative statement (that is, the basis for self-evaluation) should be deducible as well. The relative adequacy of theories employing self-acceptance constructs is clearly at issue in this case.

With regard to the problem of sampling a defined universe, one approach has been suggested by Crowne (1959). Definitions of self-accepting and self-derogatory be-

havior from the point of view of social learning theory (Rotter, 1954) were given first to psychologist judges and then to judges drawn from the subject population (introductory psychology students) to which generalization was intended. The psychologists were asked to generate from these definitions lists of self-evaluative behaviors—that is, behavioral referents, or cues, of self-acceptance and self-rejection—common in such a subject population. Subject judges were given a list of 300 adjectives (actually, the ACL) and asked to rate each adjective in terms of the extent to which they felt that, if it were checked by one of their peers as descriptive of himself, self-acceptance or self-rejection would be indicated. Items were then selected on the basis of high interjudge agreement of both psychologist and subject judges. In this way the items were tied to, and representative of, both the superordinate theory and the specific population of self-evaluative behaviors common to the experimental population. This procedure was still limited, however, to the extent that the list of 300 adjectives failed to represent some clearly defined universe. Generalizing the procedures used in this study, it would be possible to elicit descriptions of self-acceptance and self-rejection (the definitions for the judges being derived from theory) from a large sample of judges drawn from the appropriate population. Items might then be selected from descriptive units on which there was high interjudge agreement. The methodological and psychometric considerations proposed by Jessor and Hammond (1957) would presumably dictate the form of the scale, type of response, and related aspects of test construction.

This section has been concerned

with two problems related to the definition of the parameters of self-acceptance: (a) the necessity of providing a definition at the construct level, in which the behavioral referents of self-acceptance are specified and the construct located in a nomological net; and (b) the need to consider the representativeness of the sampling of a population, as defined in a of self-reference statements or items. Failure to meet these criteria results in the inability of the experimenter or test constructor legitimately to generalize from the particular conditions (subjects and stimuli) of his experiment or test.

#### SOCIAL DESIRABILITY

The third general issue to raise concerns the extent to which self-evaluative responses are influenced by "defensive behavior" (Butler & Haigh, 1954; Zuckerman & Monashkin, 1957), "self-protective response tendencies" (Crowne, 1959), or "social desirability" (Edwards, 1957; Kenny, 1956). It is important, however, first to consider whether these terms refer to the same or different phenomena.

Butler and Haigh apply the term "defensive responding" to the responses of those individuals who do not reveal the extent of their self-dissatisfaction and who, by other criteria, would be judged as maladjusted. (These authors thus seem to reject, for some subjects at least, the assumption of validity of self-reports, although how this can be done within a phenomenological frame of reference is hard to understand.) "Defensiveness" has been used by Zuckerman and Monashkin to refer to the phenomenon whereby "The person who is self-satisfied is likely to answer MMPI items in a way which he considers personally and socially desirable" (p. 147). Crowne



used the term "self-protective behavior" to refer to the unwillingness of some individuals to acknowledge self-dissatisfaction. These three terms, then, have been used to refer to highly similar phenomena.

"Social desirability" as defined by Edwards (1957), however, refers primarily to the

scale value for any personality statement such that the scale value indicates the position of the statement on the social desirability continuum (p. 3).

It also applies, as Edwards further points out,

to the tendency of all subjects to attribute to themselves, in self-description, personality statements with socially desirable scale values and to reject those with socially undesirable scale values (p. vi).

Whereas the above concepts of "defensiveness" have been applied to the motivation, presumably greater for some subjects than for others, to conceal self-dissatisfaction, Edwards' notion of "social desirability" refers to a characteristic of *items*—that is, their location on a continuum of social desirability, which determines the proportion of subjects who will attribute the characteristics to themselves.

Butler and Haigh, and also Zuckerman and Monashkin, conclude that subjects who are unwilling to attribute undesirable characteristics to themselves or confess self-dissatisfaction are by that very fact maladjusted, and presumably therefore self-dissatisfied. This, however, is obviously an hypothesis for investigation, and not necessarily true by definition. Self-acceptance tests do not directly indicate whether the subject is *willing* to express self-discontent, but only whether he *does* express it. Zuckerman and Monashkin have also suggested, in fact, that

subjects giving more socially undesirable responses may have a different conception of what is socially desirable, and thus they implicitly suggest that these subjects may actually *not* differ in terms of their *need* to respond in a socially desirable fashion. Such a difference in conception of what is socially desirable might be expected to be associated with maladjustment, but it would certainly be a less direct indication of self-dissatisfaction *per se*.

Four separate hypotheses could be advanced concerning the relationship between social desirability and responses on self-acceptance (or any other self-report) tests. Each of these is capable in some degree of being tested.

Hypothesis I. Social desirability has no effect on test responses. This is essentially the assumption of validity of self-reports: that what the subject says about himself is a valid and direct indication of what he feels or thinks, at least at the time, about himself. This, incidentally, seems to be a necessary *assumption* for phenomenologists, although it is a testable proposition.

Hypothesis II. Social desirability factors account for equal variance in all subjects' test scores. This assumption is tenable from Edwards' approach and could be held even in the face of most of the research to be reported below. It posits, in effect, that once one has accounted for variance due to nomothetically determined social desirability in any subject's test score, what is *left* indicates the subject's true self-feelings.

Hypothesis III. Social desirability, while it may or may not be an important factor for all subjects, accounts for *more* of the variance for some subjects than for others. This corresponds to the suggested differ-



ences in need to perform in a socially desirable way, protect the self, and disguise self-discontent. It is interesting that such need has been supposed to be an important variable only for those who show relatively high self-acceptance or social desirability scores: the rebel, or the individual seeking succorance, may produce very low scores, as a result of a complementary need to perform in a socially undesirable way, and still not necessarily differ from others in terms of over-all adjustment or "true" self-acceptance. In any case, such a conception as this suggests research determining the correlates of this need to perform in a socially desirable, or to perform in a socially undesirable, way.

Hypothesis IV. Variance associated with a nomothetically determined social desirability factor reflects differences in the conception of what is socially desirable. This hypothesis is not necessarily in conflict with Hypothesis III: both factors could operate simultaneously, although separating the variance due to each might be quite difficult. This, as well as Hypothesis III, is definitely incompatible with Hypotheses I and II.

With the above distinctions in mind, then, the results of some investigations of the relationship of the social desirability variable to self-acceptance test scores can be examined. Kenny (1956) gave 25 self-descriptions previously employed in a study by Zimmer (1954) to a group of judges for social desirability scaling. Three independent samples of subjects then responded to these items in the form of a questionnaire, a self-descriptive rating scale, and a *Q* sort. The correlations between the social desirability scale values and the scores obtained on the question-

naire, rating scale, and *Q* sort were .82, .81, and .66, respectively. The last two correlation coefficients are based on a "real self" scores. Social desirability correlated .82 with the "ideal self" rating scale score and .59 with the "ideal" self *Q* sort.

Edwards (1955, 1957) and Edwards and Horst (1953) have also shown that *Q* sorts are highly influenced by the social desirability variable. In a study reported in 1955 and reviewed in 1957, Edwards found correlations of .84 and .87 for males and females, respectively, between item placement on a *Q* sort and the social desirability scale values of the items. In this case, the items were those employed in the development of the Edwards Personal Preference Schedule (1953).

In one study (Kogan, Quinn, Ax, & Ripley, 1957), a social desirability scale value-response correlation of .67 was found in a hospitalized psychiatric patient sample diagnosed as psychoneurotic. The correlation in a control group of male college students was .85. It is interesting to speculate upon the possible significance of the difference in the magnitude of the correlation between self-description and social desirability values found for the patient and nonpatient groups. Perhaps, as Hypothesis IV proposes, maladjusted persons have different conceptions of social desirability in self-evaluative situations.

Studies by Berger (1955), Block and Thomas (1955), and Zuckerman and Monashkin (1957) are also relevant to the problem of social desirability. These studies investigated the relationships between self-acceptance and the clinical and "validity" scales of the MMPI. Employing different subject populations—college undergraduate students in the first two studies and hospitalized

psychiatric patients in the latter investigation—and different measures of self-acceptance, there was nevertheless considerable agreement in the findings. Self-acceptance was found to be significantly negatively correlated with a number of the clinical "adjustment" scales and positively correlated ( $r$ 's ranging from .33 to .58) with the  $K$  scale, interpreted as a measure of test-taking defensiveness (McKinley, Hathaway, & Meehl, 1948). Zuckerman and Monashkin took their findings to mean that "both self-acceptance and MMPI scales are probably being influenced more by the common trait of defensiveness than by actual adjustment" (p. 147). The term "defensiveness," with its connotation of maladjustment, seems less applicable here than "social desirability," especially in view of the high correlation (.81) reported by Edwards (1957) between the  $K$  scale and his Social Desirability Scale. With approximately 65% of the variance accounted for in the covariation of these two scales, the results of the three studies cited above would seem to be a function of the common denominator of social desirability. Thus, the items on the self-acceptance tests used and those on the MMPI are highly related to the scale values on Edwards' Social Desirability Scale.

In the study referred to earlier, Cowen and Tongas found a correlation of .91 between social desirability ratings and the self-concept score of the IAV. A correlation of .96 was obtained between social desirability ratings and the ideal-self score on the IAV. The latter correlation might be taken to suggest culturally stereotyped conceptions of what one ought to be that would be consistent with Hypothesis IV above. In another investigation (Nebergall, Angelino, & Young, 1959), it was found that sub-

jects who reported high self-acceptance tended to disagree with group judgments of adjustment. For most subjects, in fact, self-acceptance ratings were higher than group ratings. Again, these findings may be understood in terms of the individual's need to present himself in what he regards as a culturally sanctioned manner.

While this discussion has been concerned primarily with the social desirability factor in self-acceptance tests, it seems highly probable that any self-report device will be affected by the social desirability of items or of available responses. Failure to control for the effects of this variable by one of several available procedures (Edwards, 1957) means, in effect, that the test in question may better be interpreted as a measure of social desirability (that is, the subject's conception of social desirability or need to perform according to it) than of self-acceptance. This can be illustrated by means of an hypothetical experiment. It might be hypothesized that need-determined perceptual behavior—for example, perceptual reactivity to threat—is related to self-acceptance (cf. Cowen et al., 1957). Failure to control for social desirability in the self-acceptance assessment operations would make the results, no matter what the outcome, uninterpretable in terms of self-acceptance. In the light of what is already known about the influence of social desirability on self-report devices, the most probable interpretation of such an experiment would be that perceptual reactivity to threat is related (or unrelated) to the socially desirable responding of subjects—that is, their need to be perceived in a particular way or their conception of how they want to be perceived. Not provided in this experiment are the operations for deter-

mining the relationship between perceptual reactivity and "real" self-acceptance.

While studies of the effect of the social desirability variable on many of the commonly employed tests of self-acceptance have not been done, the results of the investigations discussed above would suggest that self-evaluative tests are particularly susceptible to criticism on social desirability grounds. A common denominator in research findings on self-acceptance may well be the variable of social desirability. Edwards (1957) and Jackson and Bloomberg (1958) have made a similar analysis with respect to the Taylor anxiety scale (Taylor, 1953). Systematic investigation of both the parameters and the effects on test behavior of social desirability would clearly seem to be in order. That self-acceptance tests are influenced by factors other than the manifest content of the items, however, seems beyond dispute.

#### THE GENERALITY OF SELF-ACCEPTANCE

To this point the issues discussed have been pertinent strictly to psychometric and methodological problems in assessing self-acceptance. A further issue to be raised, although it certainly has methodological ramifications, is the primarily theoretical question of the generality of self-acceptance.

Generality involves two related problems, one empirical and the other a theoretical problem of interpretation. Empirically, there is need of evidence concerning the temporal stability of self-acceptance; the consistency of an individual's self-acceptance from one situation to another (for example, in friendly vs. hostile groups, or where self-effacement is rewarded or not rewarded); the generality of self-acceptance in

reference to different aspects of the "self" (for example, in reference to morality vs. in reference to interpersonal effectiveness); and agreement of different kinds of *manifestation* of self-acceptance (for example, spontaneous self-appraisal vs. that manifested in an undisguised test such as the ACL vs. inferences drawn from a TAT protocol). The theoretical question is simply how best to construe self-acceptance. If, as has been suggested (Rogers, 1951), the self-concept and self-acceptance can be considered to be relatively stable characteristics of a person, one should find that situational variables have only a negligible effect on self-acceptance, that measures of self-acceptance taken in different social contexts are highly correlated, and that measures taken over temporal intervals are likewise highly stable. If these questions can be answered positively, it would be reasonable to construe the self-concept, from which the discrepancy notion of self-acceptance is derived, as a meaningful variable on which there are consistent differences between subjects, and it would be highly appropriate to think of individuals in terms of their characteristic levels of self-acceptance. To the degree that self-acceptance is a function of variables associated with specific situations or types of situations, however, it will be more fruitful to investigate self-evaluative behavior *per se* and its situational determinants.

The empirical evidence with respect to the generality of self-acceptance is rather scanty. The fact that studies have not attacked this question may be attributable to the general assumption that self-acceptance is consistent. Three investigations have been reported which do bear on this question. With respect to temporal stability, Taylor (1955) reports

a test-retest correlation of .79 (presumably based on self-sort-ideal-sort discrepancy scores) over an interval of a week. Butler and Haigh (1954) report the correlations between self-sorts and ideal-sorts for each subject in a control group ( $N=16$ ) not receiving therapy for two  $Q$  sort administrations separated by a considerable period of time. Although consistency was apparent, Butler and Haigh noted that

there are some sharp individual changes which indicate that alteration in self-ideal congruence does occur at times in the absence of therapy (p. 67).

Concerning the influence of situational variables of self-acceptance, a study by Thorne (1954) is relevant. Employing the IAV, Thorne found that following induced failure on a mirror drawing task, subjects whose initial level of self-acceptance was high tended to lower their self-ratings in the direction of a more realistic evaluation, while originally low self-accepting subjects tended to increase self-acceptance scores and showed concern over loss of self-esteem. The results of this study would suggest that self-acceptance is influenced by environmental events and that persons respond self-reflexively to perceived successes and failures.

It would appear, from this brief discussion, that studies should be devoted to the problem of the generality of self-evaluative behavior. Of particular interest are the questions of temporal stability, influence of situational variables, and the effect on self-evaluation of such factors as success, failure, and punishment.

#### SELF-ACCEPTANCE VS. SELF-EVALUATIVE BEHAVIOR

It has been necessary at several points in this discussion to point out

differences between a phenomenological and a behavioristic approach to self-acceptance. Since these differences are basic to the research approaches—not to mention the way in which such research is construed—in this general area of inquiry, and since these differences seem not to have been fully appreciated by all who have written on the topic, some further discussion of them is in order.

A phenomenological approach to self-acceptance is concerned with self-acceptance itself, or "real" self-acceptance, as a totally private, subjective experience of the subject. By definition this is never observable by any other; the best that an experimenter or clinician can hope to do is make relatively accurate guesses, or inferences, concerning the existence, or degree, of the variable as it "exists" in the subject. By such a definition, self-acceptance corresponds to MacCorquodale and Meehl's (1948) early conception of an "hypothetical construct"—something which cannot be observed but still is assumed to exist—except that there is little suggestion that self-acceptance even *can* be observed by anyone other than the subject himself. It is only with some difficulty, it would seem, that a phenomenologist can avoid the necessity of assuming the validity of self-reports. Representative sampling, and also an idiographic procedure for determining what are the most salient aspects of a subject's self-evaluation, would seem to be most important in a phenomenological approach to the assessment of self-acceptance. Social desirability, on the other hand, should be assumed *not* to be a factor in self-reports. To assume a high degree of generality or consistency—temporal, situational, etc.—is not necessarily essential to a phenomenological approach; however, in any theory which posits generalized self-

acceptance as an important dimension on which to compare people, empirically determined generality of the variable is, naturally, crucial.

A behavioristic concern with self-acceptance might more clearly be directed toward "self-evaluative behavior," on the other hand. The additional inference of some underlying, real if unobservable, phenomenological state is not essential to a study of self-evaluative behavior *per se*; and it might be pointed out that self-evaluative behavior is an interesting and perhaps important focus of interest in and of itself. In such an approach, the assumption of validity of self-reports is clearly not essential; a clear construct-level definition of self-evaluative behavior, on the other hand, is. Generality, representative sampling in test construction, and the related question of equivalence of assessment operations are crucial questions only if the goal is to approach self-evaluative behavior as a trait, or consistent behavioral tendency, by which to classify people in a generalized fashion. It is quite feasible to examine self-evaluative behavior as a situationally determined phenomenon, or as one determined by a situation-person interaction, rather than as a trait. Social desirability, defensiveness, etc., become merely other variables related (or unrelated) to self-evaluative behavior, and not components of error variance. And, most important, it becomes an empirical matter to determine correlates (such as "adjustment") of various forms of self-evaluative behavior, either in general or in specified contexts.

This discussion is not meant to imply that a phenomenological interest in self-acceptance is unsophisticated or unworthy. Theoretical understanding of phenomenal states is a problem of inference. A clearer

conception of "internal" phenomenal states such as self-acceptance would seem to be best derived from the observable behaviors of the person—that is, his self-evaluative behaviors. Phenomenological research would appear, in fact, to involve complexities that do not attach to more behavioristic efforts.

#### SUMMARY AND CONCLUSIONS

"Self-acceptance" promises to become an increasingly attractive focus of interest in both formal and informal psychological theory. A considerable volume of research has already been devoted to the topic, and a sizeable number of tests devised for such research. To this date, however, research has contributed an unknown, but perhaps very small, amount of understanding of self-acceptance and its relationships to other personality variables. The failures of self-acceptance research can be traced, at least in large part, to neglect of several crucial psychometric and methodological principles: the unsupported assumption of equivalence of assessment procedures, the absence of any clear construct-level definition of the variable, failure to construct tests in accord with principles of representative sampling, and questions concerning the social desirability factor in self-report tests. In addition, the absence of data concerning the generality of self-acceptance makes research results even more difficult to interpret; and the implications of the difference between a phenomenological approach to self-acceptance and a behavioristic approach to "self-evaluative behavior" have not been clearly understood.

The relative absence of systematic efforts in test development, standardization, and validation in this area is perhaps due to the fact that the focus of self-acceptance research to date has



been chiefly on the preliminary testing of hypotheses, rather than the development of adequate tests as a primary aim. A test designed solely for the purpose of testing one or two hypotheses does not, it might be argued, require so much care as a test designed to serve as a standardized instrument for many purposes. Indeed, such an argument would continue, this care and time are not usually appropriate for such restricted purposes. (The development, use, and subsequent misuse of the Taylor Manifest Anxiety Scale would serve as a case in point as Taylor herself, 1956, has pointed out.) But

when such tests are then used in further research as if they had been carefully and adequately constructed, little can ensue but error and confusion. And such seems to be the case in self-acceptance research.

Perhaps it is true that these tests are not yet used commonly in clinical settings where their inadequacies could lead to disservice to the client; perhaps it is true that the tests are used for very little other than research. But this only makes rigorous test construction the more important if research in such a complex area is to produce dependable and unambiguous results.

#### REFERENCES

- ALLPORT, G. W., & ODBERT, H. S. Trait names: A psycholexical study. *Psychol. Monogr.*, 1936, **62** (1, Whole No. 211).
- BERGER, E. M. The relation between expressed acceptance of self and expressed acceptance of others. *J. abnorm. soc. Psychol.*, 1952, **47**, 778-782.
- BERGER, E. M. Relationships among acceptance of self, acceptance of others and MMPI scores. *J. counsel. Psychol.*, 1955, **2**, 279-284.
- BILLS, R. E. Rorschach characteristics of persons scoring high and low in acceptance of self. *J. consult. Psychol.*, 1953, **17**, 36-38. (a)
- BILLS, R. E. A validation of changes in scores on the index of adjustment and values as measures of changes in emotionality. *J. consult. Psychol.*, 1953, **17**, 135-138. (b)
- BILLS, R. E. *Manual for the Index of Adjustment and Values. Form: Adult and high school senior*. Auburn: Alabama Polytechnic Institute, 1958.
- BILLS, R. E., VANCE, E. L., & McLEAN, O. S. An Index of Adjustment and Values. *J. consult. Psychol.*, 1951, **15**, 257-261.
- BLOCK, J., & THOMAS, H. Is satisfaction with self a measure of adjustment? *J. abnorm. soc. Psychol.*, 1955, **51**, 254-259.
- BROWNFAIN, J. J. Stability of the self-concept as a dimension of personality. *J. abnorm. soc. Psychol.*, 1952, **47**, 597-606.
- BUSS, A. H., & GERJUOV, H. The scaling of terms used to describe personality. *J. consult. Psychol.*, 1957, **21**, 361-369.
- BUTLER, J. M., & HAIGH, G. V. Changes in the relation between self-concepts and ideal-concepts. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954.
- CALVIN, A. D., & HOLTZMAN, W. H. Adjustment and discrepancy between self concept and inferred self. *J. consult. Psychol.*, 1953, **17**, 39-44.
- CAMPBELL, D. T., & FISKE, D. W. Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychol. Bull.*, 1959, **56**, 81-105.
- CATTELL, R. B. The description of personality: II. Basic traits resolved into clusters. *J. abnorm. soc. Psychol.*, 1943, **38**, 476-507.
- CATTELL, R. B. *Description and measurement of personality*. Yonkers-on-Hudson, New York: World Book, 1946.
- COWEN, E. L. An investigation of the relationship between two measures of self-regarding attitudes. *J. clin. Psychol.*, 1956, **12**, 156-160.
- COWEN, E. L., HEILIZER, F., & AXELROD, H. S. Self-concept conflict indicators and learning. *J. abnorm. soc. Psychol.*, 1955, **51**, 242-245.
- COWEN, E. L., HEILIZER, F., AXELROD, H. S., & ALEXANDER, S. The correlates of manifest anxiety in perceptual reactivity, rigidity, and self concept. *J. consult. Psychol.*, 1957, **21**, 405-411.
- COWEN, E. L., & TONGAS, P. N. The social desirability of trait descriptive terms: Applications to a self-concept inventory. *J. consult. Psychol.*, 1959, **23**, 361-365.
- CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281-302.
- CROWNE, D. P. The relation of self-accept-



- ance behavior to the social learning theory construct of need value. Unpublished doctoral dissertation, Purdue University, 1959.
- CROWNE, D. P., STEPHENS, M. W., & KELLY, R. The validity and equivalence of tests of self-acceptance. *J. Psychol.*, 1961, 51, 101-112.
- EDWARDS, A. L. *Manual for the Edwards Personal Preference Schedule*. New York: Psychological Corporation, 1953.
- EDWARDS, A. L. Social desirability and Q sorts. *J. consult. Psychol.*, 1955, 19, 462.
- EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.
- EDWARDS, A. L., & HORST, P. Social desirability as a variable in Q technique studies. *Educ. psychol. Measmt.*, 1953, 13, 620-625.
- GOUGH, H. G. *Reference handbook for the Gough Adjective Check-List*. Berkeley: Univer. California Institute of Personality Assessment and Research, 1955. (Mimeo)
- HILLSON, J. W., & WORCHEL, P. Self concept and defensive behavior in the maladjusted. *J. consult. Psychol.*, 1957, 21, 83-88.
- HORNEY, KAREN. *Neurosis and human growth*. New York: Norton, 1950.
- JACKSON, D. N., & BLOOMBERG, R. Anxiety: Unitas or multiplex? *J. consult. Psychol.*, 1958, 22, 225-227.
- JESSOR, R., & HAMMOND, K. R. Construct validity and the Taylor anxiety scale. *Psychol. Bull.*, 1957, 54, 161-170.
- KELLY, G. A. *The psychology of personal constructs*. New York: Norton, 1955.
- KENNY, D. T. The influence of social desirability on discrepancy measures between real self and ideal self. *J. consult. Psychol.*, 1956, 20, 315-318.
- KOGAN, W. S., QUINN, R. D., AX, A. F., & RIPLEY, H. S. Some methodological problems in the quantification of clinical assessment by Q array. *J. consult. Psychol.*, 1957, 21, 57-62.
- LAForge, R., & SUCZEK, R. F. The interpersonal dimension of personality: III. An interpersonal check list. *J. Pers.*, 1955, 24, 94-112.
- MACCORQUODALE, K., & MEEHL, P. E. On a distinction between hypothetical constructs and intervening variables. *Psychol. Rev.*, 1948, 55, 95-107.
- McKINLEY, J. C., HATHAWAY, S. R., & MEEHL, P. E. The MMPI: VI. The K scale. *J. consult. Psychol.*, 1948, 12, 20-31.
- NEBERGALL, NELDA S., ANGELINO, H., & YOUNG, H. H. A validation study of the self-activity inventory as a predictor of adjustment. *J. consult. Psychol.*, 1959, 23, 21-24.
- OMWAKE, KATHERINE T. The relation between acceptance of self and acceptance of others shown by three personality inventories. *J. consult. Psychol.*, 1954, 18, 443-446.
- PEAK, HELEN. Problems of objective observation. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences*. New York: Dryden, 1953.
- PHILLIPS, E. L. Attitudes towards self and others: A brief questionnaire report. *J. consult. Psychol.*, 1951, 15, 79-81.
- RAIMY, V. C. Self reference in counseling interviews. *J. consult. Psychol.*, 1948, 12, 153-163.
- ROGERS, C. R. *Client-centered therapy*. Boston: Houghton-Mifflin, 1951.
- ROGERS, C. R., & DYMOND, ROSALIND F. (Eds.) *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954.
- ROTTER, J. B. *Social learning and clinical psychology*. New York: Prentice-Hall, 1954.
- SARBIN, T. R., & ROSENBERG, B. G. Contributions to role-taking theory. *J. soc. Psychol.*, 1955, 42, 71-81.
- SNYGG, D. S., & COMBS, A. W. *Individual behavior: A new frame of reference for psychology*. New York: Harper, 1949.
- SULLIVAN, H. S. *The interpersonal theory of psychiatry*. New York: Norton, 1953.
- TAYLOR, D. M. Changes in the self concept without psychotherapy. *J. consult. Psychol.*, 1955, 19, 205-209.
- TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, 48, 285-290.
- TAYLOR, JANET A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, 53, 303-320.
- THORNE, R. B. The effects of experimentally induced failure on self evaluation. Unpublished doctoral dissertation, Columbia University, 1954.
- Underwood, B. J. *Psychological research*. New York: Appleton-Century-Crofts, 1957.
- ZIMMER, H. Self-acceptance and its relation to conflict. *J. consult. Psychol.*, 1954, 18, 447-449.
- ZUCKERMAN, M., BAER, N., & MONASHKIN, I. Acceptance of self, parents, and people in patients and normals. *J. clin. Psychol.*, 1956, 12, 327-332.
- ZUCKERMAN, M., & MONASHKIN, I. Self-acceptance and psychopathology. *J. consult. Psychol.*, 1957, 21, 145-148.

(Received October 30, 1959)

# THE CONSTRUCTION OF UNIDIMENSIONAL TESTS

JAMES LUMSDEN  
*University of Western Australia*

It is the purpose of this article to review methods which have been suggested, either directly or indirectly, for the construction of unidimensional tests. No general survey of this topic appears to have been made previously but much help was obtained from critiques by Loevinger (1948), Guttman (1950a, 1950b, 1950c), and White and Saltz (1957).

*Definition of unidimensional tests.* A unidimensional test may be defined simply as a test in which all items are measuring the same thing. A set of high jumps or a set of broad jumps is unidimensional. A mixture of high jumps and broad jumps is not. In psychological tests, however, items which appear to be of the same sort often turn out on closer investigation to be measuring different things so that this simple definition will not suffice for the construction of tests.

A more precise definition is given by considering the answer pattern that would be yielded by a unidimensional test with infallible items. If the items are arranged in order of difficulty placing the easiest first it will be found that a person who fails the first will fail all the other items; a person who passes the first and fails the second will fail all the subsequent items and so on. That is, the pattern of responses for five items could only be one of the forms shown in Table 1.

With fallible items where the result may be affected by fluctuations in the ability of the subjects or in the difficulty of the items a perfect answer pattern may not be found even when the items do systematically measure the same thing. For our purposes,

TABLE 1  
PATTERNS OF RESPONSES FOR A UNIDIMENSIONAL TEST OF FIVE INFALLIBLE ITEMS

Total Scores	Item				
	1	2	3	4	5
0	F*	F	F	F	F
1	P	F	F	F	F
2	P	P	F	F	F
3	P	P	P	F	F
4	P	P	P	P	F
5	P	P	P	P	P

\* P = Pass F = Fail.

however, it is sufficient to take the answer pattern of Table 1 as providing a working definition of unidimensionality remembering that with fallible items the answer pattern will be disturbed by random error.

*Criteria for evaluation.* In evaluating the methods, major consideration will be given to the extent to which a method provides for:

1. A rational procedure for item selection
2. A criterion of unidimensionality
3. An index or measure of unidimensionality

A rational procedure for item selection is essential. Any method which provides no adequate indication of the most likely items to be discarded from the pool and which relies on a blind trial-and-error procedure to discover the unidimensional set of items will be hopelessly uneconomical for practical test construction. In general the method should be convergent so that the homogeneity of the item set increases as the procedure is applied and items are pro-

gressively removed from the original pool. Minor departures from this principle at critical stages (usually the beginning) are permissible so long as the number of trials to reach a convergent state of affairs is not too large.

A criterion of unidimensionality is necessary so that checks can be made from time to time and decisions made either to continue culling of the item pool or to stop culling because a homogeneous set of items has been obtained.

An index of the closeness of approximation to unidimensionality is also required. Failure to find a set of items which meets a strict criterion of unidimensionality is certainly possible and, indeed, very likely. The set of items in question may, however, be more unidimensional than any other measuring instruments available and would be preferable to a completely heterogeneous set of items alleged to measure the same attribute. The index of unidimensionality may be related to the procedure for selecting items and/or to the proposed criterion of unidimensionality, or, on the other hand, may be quite independent of either of these. It would be desirable for the sampling distribution of the index of unidimensionality to be known.

It should be noted that the index of unidimensionality is not quite the same as the measures of reproducibility discussed by White and Saltz (1957). Reproducibility as defined by White and Saltz confounds reliability and dimensionality since the measures are affected by random errors as well as by systematic differences in item content. An index of unidimensionality appropriate to the definition used here should be independent of random error.

*Methods to be reviewed.* Explicit

consideration will be given only to classical item analysis, Loevinger's technic of homogeneous tests, the independence criterion method, Guttman's answer pattern method, and factor analysis. Most other methods are special cases of one or other of the listed methods and for the purpose of this review it is unnecessary to consider them. For example, criticisms of the Guttman procedure will apply also to the Cornell technique (Guttman, 1947a) and *H* technique (Stouffer, Borgatta, Hays, & Henry, 1952). Certain related techniques such as the Thurstone attitude scaling methods give tests of unidimensionality as a by-product but as test construction methods they are subject to the same criticisms as classical item analysis and the independence criterion.

#### CLASSICAL ITEM ANALYSIS

Classical item analysis using an internal criterion attempts among other things to increase the average item-test correlation by selecting from the item pool those items which have the highest item-test correlation. It is well-known that this procedure tends to increase the homogeneity of the test.

From Table 1 it will be clear that for infallible items forming a unidimensional test the item-test correlation will be the maximum permitted by the shape of the distribution of test scores. With the answer pattern of Table 1 there is no overlap in the distribution of test scores for those who pass and those who fail a given item. The difference in mean test scores of passers and failers is thus a maximum and the biserial correlation between item and test is consequently maximized. It would appear then that if the culling of items proceeds to the point where the item-test

correlations are all maximized the resulting test would be unidimensional. There are a number of difficulties which make this program unlikely to succeed.

With fallible items the maximum item-test biserial will not be reached. One solution would be to correct the obtained biserials for attenuation using estimates of the reliability of item and test scores. Accurate estimates of the reliability of a single item are not easily obtainable. Assuming that this difficulty can be overcome a test would be regarded as unidimensional if the biserial correlations between item and test approached the maximum after correction for attenuation.

Even granted the assumption that accurate estimates of item reliabilities can be obtained the method is not satisfactory. Consider the set of items with factor constitutions as follows:

$$x_1 = ma + nb + e_1$$

$$x_2 = ma + nb + e_2$$

$$x_3 = ma + nb + pc + e_3$$

$$x_4 = ma + nb + qc + e_4$$

$$x_5 = ma + nb + rc + e_5$$

where  $a$ ,  $b$ , and  $c$  represent different orthogonal common factors,  $m$ ,  $n$ , and  $p$ ,  $q$ ,  $r$  are loadings; and  $e_1$ ,  $e_2$ ,  $e_3$ ,  $e_4$ , and  $e_5$  are error factors.

Lumsden (1957) has shown that Items 1 and 2 form a unidimensional subtest and that Items 3, 4, and 5 with differing loadings on  $c$  are not unidimensional. Yet the method of maximizing item-test correlations will eliminate Items 1 and 2 first and no unidimensional test will be discovered. The only way out of this impasse would be to try sets at random which would make the procedure nonrational.

For this method the criterion of unidimensionality would be maximum biserial after correction for attenuation. No sampling distribution of corrected biserials appears to be available so that the significance of departures from the perfect fit cannot be assessed. This is specially important in this case since the estimates of item reliability on which correction is based are likely themselves to be quite unreliable.

The logical measure of unidimensionality would be average corrected biserial. This would need to be considered relative to the maximum obtainable biserial (biserial  $r$  has a maximum of 1.0 only when the continuous variable is normally distributed). A ratio of corrected biserial to its maximum similar to Loewinger's  $H_1$  suggests itself but the absence of a knowledge of its sampling distribution would restrict its value.

An obvious possibility would be to use the Kuder-Richardson Formula 20 with correction for variation in item difficulty suggested by Horst (1953). This statistic is, however, affected by random as well as systematic variance and is therefore, a measure of reproducibility rather than an index of unidimensionality. There would seem nothing to prevent the development of an index based on the ratio of obtained K-R 20 to the maximum K-R 20 for items with a given amount of random error.<sup>1</sup>

A search of the literature has not revealed any writer who has advocated the use of classical item analysis techniques as described above in order to produce unidimensional tests. Thorndike attempted to demonstrate the "homogeneity of intellect CAVD" by correlating scores on subgroups of

<sup>1</sup> I am indebted to John Ross (University of Sydney) for this suggestion.

items with scores on the total set of items and correcting the obtained  $r$ 's for attenuation. Evidence was presented (Thorndike, Bergman, Cobb, Woodyard, 1926, p. 566) that these corrected correlations approximated 1.0 and Thorndike concluded that this demonstrated the homogeneity of CAVD tests. The logic of Thorndike's procedure is impeccable if applied to single items or to randomly selected subgroups of items but his subgroups were arranged so as to have, like the total set, equal numbers of Completion, Arithmetic, Vocabulary, and Directions items. Thorndike was thus merely able to show that the composite score ob-

#### LOEVINGER'S TECHNIQUE OF HOMOGENEOUS TESTS

Loevinger's procedure is closely related to classical item analysis and indeed she indicates (Loevinger 1947, p. 26) that the earlier work by Thorndike on the CAVD tests may have been influential in the development of her procedure.

The procedure is based on two statistics: the "homogeneity of an item with a test" and the "homogeneity of a test." The first of these is to be used as a tool for item selection and is a development of Long's (1934) index of overlapping. The formula for this is given by Loevinger as:

$$H_{ii} = 1 - \frac{2 \text{ ("passes" below or tied with "fails")}}{PQ - \text{"passes" one above "fails"}}$$

tained from his subsets was similar to the total score obtained from the complete set but not that the subsets or the total set were homogeneous in the sense used here. It is only fair to point out that Thorndike was mainly concerned to show that his easier sets of items and his harder sets gave the same sort of results as the total set.

Wherry and Gaylord (1943) suggest as an alternative to factor analysis an iterative procedure based on classical item analysis. In this procedure each item is correlated with total score; those items with the highest correlations are selected and a new total formed; all items (including those not selected in the first stage) are then correlated with the new total and the procedure is continued until a stable group of items is obtained. White and Saltz (1957) commend this method but it would not appear to avoid any of the difficulties of classical item analysis.

where  $P$  is the number passing the item and  $Q$  is the number failing the item. It is clear that for a perfectly unidimensional test as defined by Table 1  $H_{ii}$  will equal 1.0 since there will be no subjects who pass an item who will have scores below or tied with subjects who fail the item. Using this statistic to cull a mixed set of items will, however, be subject to all the difficulties encountered with classical item analysis.

The index of unidimensionality is provided by the "homogeneity of a test,"  $H_t$ . Loevinger notes that for a perfectly heterogeneous test  $p_{i/j} = p_i$  (i.e., probability of passing an Item  $i$  having passed another Item  $j$  is the same as the overall probability of passing Item  $i$ ). For a perfectly homogeneous test as defined by Table 1,  $p_{i/j} = 1.0$  for all cases where  $p_i > p_j$  (i.e., where Item  $i$  is easier than Item  $j$ ). From this it will be seen that  $p_{i/j}$  has a minimum value of  $p_i$  for all cases where  $p_i > p_j$ .

Loevinger then considers the sum:

$$S = \sum_{i=1}^{m-1} \sum_{j=i+1}^m p_j(p_{i/j} - p_i)$$

where  $m$  is the number of items and the item pairs are all such that  $p_i > p_j$ .

This sum will have a maximum value given by

$$S_{\max} = \sum_{i=1}^{m-1} \sum_{j=i+1}^m p_j(1 - p_i)$$

for a perfectly homogeneous test and a value of zero for a perfectly heterogeneous test. To provide an index with the formal properties of a minimum of zero and a maximum of 1.0, Loevinger divides  $S$  by  $S_{\max}$  to give:

$$\frac{S}{S_{\max}} = \frac{\sum_{i=1}^{m-1} \sum_{j=i+1}^m p_j(p_{i/j} - p_i)}{\sum_{i=1}^{m-1} \sum_{j=i+1}^m p_j(1 - p_i)}$$

Loevinger provides a formula for estimating  $H_t$  from sample statistics but points out that the sampling distribution is unknown and that the estimate is not even known to be unbiased.

#### INTERDEPENDENCE CRITERION

Lazarsfeld (1950), Tucker (1952), and Lord (1952) have pointed out that with a unidimensional test the probability of success on one item is independent of success in any other item for subjects with the same true score. This is at first sight paradoxical because it would seem obvious that items which are measuring the same thing should be highly correlated. But when only subjects of the same true ability are considered then items which are measuring this ability and nothing else can differ only through error and will exhibit no

systematic variance. If we take subjects who are exactly 6 feet tall then different measures of height will vary only through error so that the measurements will be independent, uncorrelated. The independence criterion is undoubtedly valid and is more general than any other. It makes no assumptions about the distribution of ability or rectilinearity of regression.

The criterion suggests a procedure for constructing unidimensional tests. It would be possible to obtain results on a pool of items from a large group of subjects, to choose a number of subjects with the same total score, and then to determine say by  $\chi^2$  whether the items are independent or not. If certain items turned out not to be independent these could be rejected, new totals worked for all subjects in the original group, a new group with the same total score determined, and the  $\chi^2$  test repeated.

The true scores on the test are not, however, known and the estimates obtained from the raw test scores are not satisfactory. O'Neil (1954) has shown that, for subjects with the same obtained score, items, even if unidimensional, tend not to be independent but to be negatively correlated. If there are only two items for example then for subjects with an obtained score of 1 the items have a tetrachoric correlation of  $-1.0$  since if a subject has the first item right he must have the second one wrong and vice versa. In mathematical terms  $p_{i/j} = 0$  instead of  $p_i$  as required by the independence criterion. This effect is known to decrease as the number of items is increased and it is possible that the independence criterion may be workable for fairly large groups of items. With infallible items there is, of course, no problem since true scores will then equal the ob-



tained scores and the quoted example could not occur (if the items were unidimensional).

Even if this problem is overcome, the culling of items is likely to prove arduous. All items in the pool are likely to be correlated on the first trial. In the absence of any knowledge about the number of items in the unidimensional set it is impossible to say whether the unidimensional items will be more or less intercorrelated than the items it is desired to reject. No rational, convergent procedure of item culling is available using the independence criterion.

No special index of unidimensionality is suggested for this method. This, of course does not matter, since if the method was otherwise suitable an index could be borrowed from one of the other methods.

#### ANSWER PATTERN METHODS

The Guttman procedure (1944) is the most important of the answer pattern methods and is the only one discussed here. Some earlier writings by Walker (1931, 1936, 1940) and Ferguson (1941) have the first explicit discussions of the relationship between answer pattern and other test characteristics but no suggestions for test construction were made.

The answer pattern procedure consists essentially of inspecting the answer pattern and removing items so that the remaining items have patterns which are as near as possible to those of Table 1. It is clear that for infallible items this procedure could be easily carried out and that a simple inspection of the answer patterns would provide a clearcut criterion of unidimensionality. For items which exhibit slight departures from unidimensionality the procedure would be to eliminate items until the closest possible approximation con-

sistent with retaining a sufficient number of items was obtained. For this, some measure of the closeness of approximation to unidimensionality is required. Guttman uses the coefficient of reproducibility which is the proportion of responses which can be correctly predicted from the total raw score. For a perfectly unidimensional test it will be seen from Table 1 that the reproducibility coefficient will have the value 1.0. Guttman suggests that a test may be regarded as a "scale" (i.e., as unidimensional) if the coefficient of reproducibility exceeds .90.

The coefficient of reproducibility has been criticised severely by Festinger (1947) and Jackson (1949) because it does not allow for the chances of obtaining high values when the items are heterogeneous (e.g., with only a few items of widely differing difficulties). Guttman (1947b) replied to criticism claiming that such factors as the number of answer categories and the range of difficulty were taken into account before calculating the coefficient of reproducibility. Guttman does not give explicit rules but improvements to the reproducibility coefficient have been suggested by Jackson (1949) and Green (1954) which overcome some of the problems.

The reproducibility coefficient, however modified, does not permit of a distinction being made between random and systematic scale discrepancies. Guttman claims (1950a, 1950b, 1950c) that the distinction may be made by examining the patterns of scale discrepancies and presents tables (p. 161) which purport to represent scale patterns for a perfect scale, a scale with random error, and a scale with systematic error. Evidence for random error in an item is said to be provided when scale errors are distributed randomly around the

cutting point for the item; evidence for systematic error when the scale errors are grouped in a systematic fashion. While this claim is undoubtedly correct (such systematic groupings are the basis of all the statistical analyses proposed for the problem) it is difficult to see how these groupings may be discovered by inspection and distinguished from random errors when the random errors are fairly large.

Guttman (1950b) has explicitly denied any intention to use scale analysis for the selection of items. His scalogram was designed merely to discover approximate cutting points for attitude scale items. Guttman indeed claims that the task of scale analysis is to discover scales rather than to construct them and states that if a universe of attributes is scalable then any subset of items from that universe is scalable. Item culling is by this argument unnecessary. The difficulty is that a test constructor (or discoverer) does not know precisely what "universe of attributes" he is sampling. Without precise definition he may sample a number of related universes. Item culling procedures are designed to distinguish between groups of items selected from different universes.

It may be seen then that the answer pattern method provides no rational culling plan for use with fallible items. The index of unidimensionality provided by the plan is the coefficient of reproducibility which, despite improvements on the early Guttman form, does not distinguish between systematic and random error.

#### FACTOR ANALYSIS

It is difficult to give due credit to whoever first suggested the use of factor analysis in the construction of

unidimensional tests. The idea is sufficiently obvious to be thought at least implicit in the writings of Spearman, Thurstone (1947), and other early factorists. The factor analyses of test items by McNemar (1942), Burt and John (1943), and others clearly suggest it. Papers on related topics by Ferguson (1941), Wherry and Gaylord (1943, 1944), Carroll (1945), and Loevinger (1948) discuss with varying degrees of completeness the possibility of factor analyzing items in test construction.

Under restrictions which appear plausible for ability test items it is easy to show (vide Lumsden, 1957) that for a unidimensional test the matrix of tetrachoric item intercorrelations is of unit rank. One factor analytic procedure for constructing unidimensional tests is to extract a single factor from the item intercorrelations, cull out the items which have large residuals, reanalyze, and continue until a satisfactory fit to a single factor solution is obtained. Woffle (1940) in a well-known jibe at Brown and Stephenson (1933) said: "if one removes all tetrad differences which do not satisfy the criterion, the remaining ones do satisfy it" (p. 9). That is exactly what is done in this factor analytic technique of constructing unidimensional tests. The difference between the two situations is, of course, that Brown and Stephenson had asserted that their tests, all of them, would meet the tetrad difference criterion, while here it is merely hoped that a subset of items will meet the criterion.

The procedure is quite simple. But is the culling procedure rational? Will the set of items converge to unidimensionality?

It is evident that convergence of the factor analytic procedure to a unidimensional subset of items can-

not be guaranteed. If the unidimensional set is much less numerous than the heterogeneous items in the pool then it is probable that the unidimensional set will not have sufficient influence on the nature of the first factor extracted to prevent the occurrence of large residuals among the unidimensional set. These items will be discarded first and the procedure will not converge to a single factor solution.

If, however, the items are carefully preselected on empirical and *a priori* grounds, it seems likely that the state of affairs of the preceding paragraph will not occur. If items are deliberately made parallel or if there is evidence for parallelism then it would follow that the dimension of any unidimensional test and the dimensions of the heterogeneous items in the total pool, will normally be highly correlated. In this circumstance the influence of the unidimensional set on the first factor extracted may well be greater than the actual numbers of items suggest, and the method may therefore be expected to converge. The procedure of preselecting will also tend to increase the size of the unidimensional set in the pool and this will also increase the probability of convergence.

Lumsden (1959) found that four subsets of number series items selected on *a priori* grounds converged rapidly and that three of them met a fairly stringent test of unidimensionality when cross-validated with a fresh group of subjects.

One procedure that should almost guarantee convergence (if a sizable unidimensional set exists) is to carry out a preliminary complete centroid analysis and then to select for further analysis those items which appear in narrow strips (i.e., roughly co-linear) in the factor space. This appears to

be the procedure advocated by Cattell (1957) for his factor homogeneous scale except that he would require the additional restriction that the factor have significance in a more general factor space than that provided by the item intercorrelations. The complete centroid procedure with rotation could indeed be used without further analysis except that the problems of estimating communalities and determining goodness of fit are more complicated than for the unit rank case.

The criterion of unidimensionality suggested for item culling is the size of the residuals. This must be considered with relation to the sampling distribution of residuals. Unfortunately there is no exact solution to this problem. Many methods have been suggested (Cattell, 1952) but none can be regarded as satisfactory. A reasonable solution for test construction purposes would be to use one of the simpler procedures (e.g., standard error of average  $r$ ) and apply it rather severely. Increased availability of automatic computing services may permit the use of maximum likelihood methods of factorizing which provide a test for rank.

An index of unidimensionality appropriate to the method is the ratio of first factor variance to total bipolar factor variance after a complete centroid analysis with subjects who were not used for item selection. In most cases the ratio of first to second factor variance would seem to give a reasonably useful index. This index has no fixed maximum value and little is known about the extent to which it may be affected by errors of sampling or of measurement.

#### DISCUSSION

It seems clear that none of the methods examined can be regarded as

satisfying all three of the main criteria. Only factor analysis appears to offer a rational procedure for item selection. The criteria and indices of unidimensionality are unsatisfactory for all methods.

This review has considered each of the methods as if they were complete, self-consistent creations of a single writer. With the exception of the Guttman answer pattern method and the Loevinger method this is not so. The various "natural" criteria and indices suggested for each of the methods are not necessary consequences of the choice of item selection method. Combinations of different elements from different methods are possible and this circumstance justifies a modified optimism. Thus a modification of the coefficient of reproducibility which produced an acceptable index of unidimensionality would not be cogent evidence for adopting an answer pattern method but would greatly improve all methods.

Greatest emphasis has been deliberately placed on item selection rationale since this topic appears to have been relatively neglected in the

literature of the problem. Great advances appear unlikely unless the development of criteria and indices of unidimensionality is closely related to item selection procedures.

#### SUMMARY

Five methods of constructing unidimensional tests (classical item analysis, Loevinger's procedure, the independence criterion method, the answer pattern method, and factor analysis) have been considered with respect to their provision for: a rational procedure for item selection, a criterion of unidimensionality, and an index of unidimensionality.

It has been argued that only factor analysis provides a rational procedure for item selection. No method has a fully satisfactory criterion of unidimensionality. The index of unidimensionality suggested for the factor analytic method is the ratio of first to second factor variance. This suffers from the absence of any knowledge of sampling fluctuations, but this weakness is shared by the only reasonable alternative, the coefficient of reproducibility.

#### REFERENCES

- BROWN, W., & STEPHENSON, W. A test of the theory of two factors. *Brit. J. Psychol.*, 1933, 23, 352-370.
- BURT, C., & JOHN, E. A factorial analysis of the Terman-Binet tests. *Brit. J. educ. Psychol.*, 1943, 12, 156-161.
- CARROLL, J. B. The effect of difficulty and chance success on correlations between items or between tests. *Psychometrika*, 1945, 10, 1-19.
- CATTELL, R. B. *Factor analysis*. New York: Harper, 1952.
- CATTELL, R. B. *Personality and motivation structure and measurement*. New York: World Book, 1957.
- FERGUSON, G. A. *The reliability of mental tests*. Univer. London Press, 1941.
- FESTINGER, L. The treatment of qualitative data by "scale analysis." *Psychol. Bull.*, 1947, 44, 146-161.
- GREEN, B. F. Attitude measurement. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954.
- GUTTMAN, L. A basis for scaling qualitative data. *Amer. sociol. Rev.*, 1944, 80, 139-150.
- GUTTMAN, L. The Cornell technique for scale and intensity analysis. *Educ. psychol. Measmt.* 1947, 7, 247-279. (a)
- GUTTMAN, L. On Festinger's evaluations of scale analysis. *Psychol. Bull.*, 1947, 44, 451-465. (b)
- GUTTMAN, L. The basis for scalogram analysis. In S. A. Stouffer (Ed.), *Measurement and prediction*. Princeton: Princeton Univer. Press, 1950. (a)

- GUTTMAN, L. The problem of attitude and opinion measurement. In S. A. Stouffer (Ed.), *Measurement and prediction*. Princeton: Princeton Univer. Press, 1950. (b)
- GUTTMAN, L. Relation of scalogram analysis to other techniques. In S. A. Stouffer (Ed.), *Measurement and prediction*. Princeton: Princeton Univer. Press, 1950. (c)
- HORST, P. Correcting the Kuder-Richardson reliability for dispersion of item difficulties. *Psychol. Bull.*, 1953, **50**, 371-374.
- JACKSON, J. M. A simple and more rigorous technique for scale analysis. In, *A manual of scale analysis*. McGill University, 1949. (Mimeo)
- LAZARSFELD, P. F. The logic and mathematical foundation of latent structure analysis. In S. A. Stouffer (Ed.), *Measurement and prediction*. Princeton: Princeton Univer. Press, 1950.
- LOEVINGER, JANE. A systematic approach to the construction and evaluation of tests of ability. *Psychol. Monogr.* 1947, **61**(4, Whole No. 285).
- LOEVINGER, JANE. The technic of homogeneous tests compared with some aspects of "scale analysis" and factor analysis. *Psychol. Bull.*, 1948, **45**, 507-529.
- LONG, J. A. Improved overlapping methods for determining the validities of test items. *J. exp. Educ.*, 1934, **2**, 262-268.
- LORD, F. M. A theory of test scores. *Psychometric Monogr.*, 1952, No. 7.
- LUMSDEN, J. A factorial approach to unidimensionality. *Aust. J. Psychol.*, 1957, **9**, 105-111.
- LUMSDEN, J. The construction of unidimensional tests. Unpublished master's thesis, University of Western Australia, 1959.
- MCNEMAR, Q. *The revision of the Stanford-Binet scale*. Boston: Houghton Mifflin, 1942.
- O'NEIL, W. M. A problem of testing a set of items for unidimensionality. Paper read at British Psychological Society (Australian Branch), Perth, 1954.
- STOUFFER, S. A., BORGATTA, E. F., HAYS, D. G., & HENRY, A. F. A technique for improving cumulative scores. *Public opin. Quart.*, 1952, **16**, 273-291.
- THORNDIKE, E. L., BERGMAN, E. D., COBB, M. V., & WOODYARD, E. *The measurement of intelligence*. New York: Teachers' College, Columbia Univer., 1926.
- THURSTONE, L. L. *Multiple factor analysis*. Chicago: Univer. Chicago Press, 1947.
- TUCKER, L. R. A level of proficiency scale for a unidimensional skill. *Amer. Psychologist*, 1952, **7**, 408. (Abstract)
- WALKER, D. A. Answer-pattern and score scatter in tests and examinations. *Brit. J. Psychol.*, 1931, **22**, 73-86.
- WALKER, D. A. Answer-pattern and score scatter in tests and examinations. *Brit. J. Psychol.*, 1936, **26**, 301-308.
- WALKER, D. A. Answer-pattern and score scatter in tests and examinations. *Brit. J. Psychol.*, 1940, **30**, 248-260.
- WHERRY, R. J., & GAYLORD, R. H. The concept of test and item reliability in relation to factor pattern. *Psychometrika*, 1943, **8**, 247-269.
- WHERRY, R. J., & GAYLORD, R. H. Factor pattern of test items as a function of the correlation coefficient: Content, difficulty and constant error factors. *Psychometrika*, 1944, **9**, 237-244.
- WHITE, B. W., & SALTZ, E. Measurement of reproducibility. *Psychol. Bull.*, 1957, **54**, 81-99.
- WOLFLE, D. Factor analysis to 1940. *Psychometric Monogr.*, 1940, No. 3.

(Received January 4, 1960)

## CONTRIBUTIONS AND CURRENT STATUS OF Q METHODOLOGY

J. R. WITTENBORN  
*Rutgers University*

For the present purposes we shall take our definition of *Q* methodology from Stephenson (1953), not because he has been succinct, but because he has been more comprehensive in his interest than other writers. Accordingly, *Q* as conceived by Cattell or Mowrer and related methodological topics discussed by such writers as Cronbach are not included in the present review of recent studies.

In his 1953 publication, *The Study of Behavior*, Stephenson informs us with a modesty which is characteristic for this book that "... the science of behavior can be immeasurably improved by attending to a few principles upon which we have based the method now well known as 'Q-technique' " (p. 1). Time does not permit the long series of quotations which would be necessary in order to indicate fully Stephenson's concept of *Q* methodology, but a few additional quotations remind us that he was definite in his point of view.

Our object has been to make it possible for studies to be undertaken on single cases (p. 2).<sup>1</sup> Briefly, a statement of the kind "All crows are black" is a general proposition. To say that "A crow is black" is clearly singular, but not testable. When, however, we can point to a particular crow *X* and assert that *it* is black, a singular testable proposition is at issue (p. 42). There never was a single matrix of scores to which both *R* and *Q* apply (p. 15).

<sup>1</sup> "By a 'single case' we mean, for the moment, a single person under study or a single group of interacting persons. . . . what is involved is whether individual differences are postulated or whether singular propositions are being tested. The latter alone are our concern" (Stephenson, 1953, p. 2).

A defining summary of Stephenson's view of *Q* methodology could include at least six points:

1. *Q* method appears to require ipsative variables, particularly *Q* sorts.
2. *Q* method lends itself to correlations between people or between different conditions for the same person.
3. *Q* method requires a conceptually structured set of statements in order to interpret the correlations between people—each set of statements comprising systematic combinations of different levels of the various hypothetical effects.
4. *Q* method permits a study of a person by means of analysis of variance of the statements, assuming that the sorted statements were initially structured as replications of the possible combinations of *a priori* effects and levels of reaction.
5. *Q* method favors a dependency type emphasis in factor analyses with rotations determined by the nature of the propositions concerning the variates.
6. *Q* method leaves unanswered the question of the parent population from which the individual is drawn; the method examines singular propositions on the assumption that somewhere there are more people like the one under scrutiny.

To date the most conspicuous use of *Q* methodology has been made by the so-called "self psychologists" who view discrepancies between one's self-perception and the perception of an ideal self as an indication of



maladjustment. This interpretation of psychological maladjustment is consistent with Rogers' belief in the self-actualizing function of the personality and, as a consequence, finds direct application in his studies of the efficacy of psychotherapy (Rogers & Dymond, 1954). The work done by the Chicago group during the first half of the present decade used the now familiar device of correlating the individual's *Q* sort describing his true self with his *Q* sort for his ideal self. Increases in these correlations during the course of therapy were taken as evidence of improvement. This was a notable application of *Q* methodology because in the opinion of many persons these studies comprised the first acceptable indication that psychotherapy was efficacious in producing personality change. Beyond using *Q* methodology in establishing this important landmark, the Chicago group was able to illuminate some of the features of the psychotherapeutic process by factor analyzing the intercorrelations among various *Q* sorts for a given patient. The case of Mrs. Oaks is illustrative. During the course of therapy her concept of self became much more favorable, and there were some changes in her concept of an ideal self. Her therapeutic progress was summarized by a factor analysis of the intercorrelations among *Q* sorts made at various stages in the course of her therapy.

There have been several evaluations of the validity of *Q* sorts as evidence of adjustment, particularly from the standpoint of their appropriateness as criteria for therapeutic change. For example, Friedman (1955) supported the self-ideal consistency concept of good adjustment by a study which involved a comparison between normals and neu-

rotics. The neurotic group was described as tending to regard their self-qualities as very much different from the way they would like them to be. Cartwright (1957) emphasized the consistency interpretation of good adjustment by showing that after psychotherapy subjects describe themselves in relation to important persons in their environment with as much consistency as controls. An increase in self-ideal congruence for a group of high school boys after counseling was reported by Caplan (1957), and Turner and Vanderlippe (1958) reported that college students with high self-ideal congruence tended to have more extracurricular activities and to have higher scholastic averages than students with low self-ideal congruence.

The apparent validity of self-ideal congruence was examined by Chase (1957), who compared adjusted and maladjusted hospital cases with respect to the various possible correlations involving *Q* sorts. Only those correlations containing the self-sort distinguished between the adjusted and the maladjusted group.

The *Q* sort approach to adjustment is subjected to further scrutiny by Kogan, Quinn, Ax, and Ripley (1957). Using two comparable samples, one psychiatric patients and the other university students, sorts for four different conditions were obtained. The average sorts for the patient and the student groups were correlated for each of the four conditions. It was found that a great portion of the variance in these correlations could be accounted for in terms of either of two extraneous variables, the social desirability of the sorted statements or a sickness-health variable. Edwards (1955) had described the importance of social desirability in *Q* sorts as early as 1955.

There are reports which challenge the appropriateness of *Q* sorts as a direct evidence of the efficacy of psychotherapy. For example, Taylor (1955) undertook a study on the assumption that repeated introspection would produce therapeutic type changes in self-concept. His subjects made repeated *Q* sorts. In consequence of this procedure, there was an increase in positive self-concepts, and the self-ideal correlation increased. From this one need not infer that self-concepts are unstable, however. Engel (1959) examined the self-concepts of a group of adolescents in 1954 and again in 1956, and reported that items relative to positive self-concepts had appreciable stability as indicated by a stability correlation of .69.

Levy (1956) challenged the meaning of self-ideal discrepancies by comparing self-ideal correlations based on the Butler and Haigh (1954) items with the correlation between sorts for an actual and an ideal home town. He found these two sets of actual-ideal correlations to be correlated with each other to the order of .70. Because of this, he suspects that the discrepancies perceived between actual and ideal states of affairs have implications for the individual's adjustment, regardless of the area in which the discrepancy is shown.

Although *Q* sorts are frequently employed in the published literature, the use is often relatively uncritical. In some instances it would appear that a normative procedure would have served the investigator's purposes better than the ipsative *Q* sorts. It appears probable, however, that investigators have been encouraged by the availability of *Q* sort procedures, and some of the resulting studies might not have been undertaken if only a normative type em-

phasis had been available. For example, in Stewart's (1958) study of the relationship between manifest anxiety and mother-son identification, facets of mother-son identification are readily revealed and usefully quantified by correlations between various sorts provided by the mothers and their sons. Stewart found that the boys with the greatest manifest anxiety were those with the greatest discrepancy between their self-perceptions and their mother's ideal for them.

Correlating *Q* sorts was a convenient device for Kalis and Bennett (1957), who wished to show that communication between the patient and members of his family was improved for those patients whose hospitalization had been successful. The importance of similarity of self-perceptions in interpersonal relationships is further illuminated by Corsini's (1956) use of *Q* sorts in his study of happiness in marriage. These studies are reminiscent of a report by Revie (1956), who used *Q* sorts to describe both the teacher's and the school psychologist's concept of pupils. As a result of their interaction, both the teacher's and the psychologist's concept of the pupil changed.

*Q* sorts have been used in many different ways, particularly in the study of personality. Shontz (1956) used *Q* sorts in order to examine the concept of a healthy personality, while Reznikoff and Toomey (1958) worked out a system of weightings whereby observers' *Q* sorts of patients may be scored to estimate the degree of emotional disturbance. Epstein and Smith (1956) used *Q* sorts as a sociometric device by having students *Q* sort their fellows with respect to the degree of hostility in their behavior. Fiske and Van Bus-

kirk (1959) used *Q*-sort procedures in order to examine the stability of sentence completion test interpretations, and Doleys and Kregarman (1959) report that self-ideal congruence does not measure frustration tolerance. Nahinsky (1958) used a self-ideal comparison to distinguish career from noncareer naval officers, and Whiting (1959) had nurses, patients, and physicians sort statements concerning the importance of various aspects of the nurse's work. This appears to be the kind of a study where rating scales, inventories, or check lists could not have served the investigator's purposes as well as the ipsative sort.

The unique value of *Q* sorts has not been made sufficiently explicit to permit an investigator to know the kinds of situations which call for ipsative procedures and the kinds of situations in which his purposes will be better served by normative procedures. There are numerous studies in the literature which employ *Q* sorts without indicating why this particular method was chosen. Sometimes it appears that *Q* sorts are used because no reliable normative instrument is available to distinguish between persons along a relevant continuum. The question of the reliability or the validity of the *Q* sort is rarely raised, and if practice alone were considered, one could infer that reliable and valid ipsative distinctions based on a *Q*-sort procedure are much easier to establish than reliable and valid normative procedures. Even if this were true, and your reviewer has not seen material which would lead to this conclusion, one could still question the use of an ipsative procedure showing intra-individual differences when a normative procedure dealing with inter-individual differences appears to be

indicated by the general requirements of the investigation. For example, Morsh (1955) described the use of a *Q*-sort procedure in securing the classes' evaluation of the teachers. Why an ipsative type evaluation of teachers is preferable to a normative-type procedure is not indicated in his report.

In a study of the relationship between some personality variables and speed in decision making, Block and Peterson (1955) used the staff's *Q* sort of the subject as a measure of personality. Although the results of this investigation are interesting and worthwhile, it would appear that the emphasis is one of individual differences and that a normative measure of personality would have been the logical choice. Both Cattell (1944) and Guilford (1954) have warned that ipsative measures should not be used in attempts to study individual differences. The amount of error in such a maneuver need not be invariably great, however. For example, Block (1957) matched items from a *Q* sort with items that were used in a normative-type rating and found that in one sample the correlations between various items ranged from .63 to .88, while similar correlations for another sample ranged from .31 to .74. Apparently, the error involved in using ipsative item scores in a normative manner may vary greatly from item to item and from sample to sample.

In addition to their applications in various studies of personality, *Q* sorts are also applied in the study of psychopathology. For example, Rogers (1958) found that the self-ideal congruence for paranoid schizophrenics was greater than for normals. His approach is noteworthy because of its novelty. Instead of having his subjects sort cards, he

asked them to manipulate a red square over a blue square, with the red square representing the self, the blue square representing the ideal, and the overlap representing the degree of congruence. Although the spatial interpretation that the subject gives his judgment is absolute and could lend itself to normative treatment, the sense of these manipulations is clearly ipsative. This study, which was published in 1958, showed a high degree of self-ideal congruence for paranoids and can be compared with Friedman's 1955 study which included a sample of paranoid schizophrenics. Friedman found that only 3 of his 16 paranoids showed a low self-ideal correlation.

Other schizophrenics are much more distinctive with respect to their behavior in the *Q*-sort situation. For example, Helfand (1956) asked various subjects, including schizophrenics, to simulate the *Q* sort of a former patient whose autobiography they read. He then computed the correlations between the sorts provided by his various subjects and the sort provided by the former patient. He found that the schizophrenics' simulated sorts had the lowest correlations of all. He ascribes this to a limitation in role-taking ability. A recent paper by Fagan and Guthrie (1959) tells us more about the schizophrenics. Their subjects were asked to describe themselves in one sort and to describe an average person in another. The subjects were intercorrelated for the two different sorts, and the two sets of intercorrelations were factor analyzed. The authors concluded that schizophrenics, like many other patients, view themselves differently from the way they view other persons.

The mothers of schizophrenic patients have also been studied by *Q*

methodologists. Shepherd and Guthrie (1959) had the mothers of 20 male schizophrenics sort 100 items concerning children and family life. These sorts made it possible to correlate each mother with every other one. The correlations were factor analyzed and the resulting factors—identified as Detached Authoritarianism, Inadequacy and Inconsistency, Pervasive Control, Sophisticated Denial of Inadequate Mothering, and Annoyance and Rejection—broaden our view of the various qualities or dimensions of schizophrenic mothering. One immediately becomes interested in the manner in which one might generalize from these mothers of schizophrenics to other mothers and thereby gauge how broadly applicable one might find such dimensions of schizophrenic mothering. Unfortunately, this is one of the ways in which *Q* methodology is weakest. We do not know what population the individual or individuals under scrutiny represent. Stephenson (1953) seems to feel that this really does not matter as long as he can assume that there are other similar individuals somewhere. He calls ducking this practical issue testing a "singular proposition."

There are many published studies which involve factor analyzing intercorrelations between persons. According to Stephenson's criteria, however, only a few of these would qualify as an application of *Q* methodology. Since Stephenson states that there is no matrix of correlations which can be studied by both *R* and *Q* methods, one is inclined to conclude that one can appropriately intercorrelate persons for *Q* purposes only when the similarity of the persons is expressed by a correlation based on ipsative scales, i.e., scales on which people have distinguished

between items and not necessarily scales which distinguish between people on any normative basis of individual differences. Thus, your reviewer's factor analyses of various diagnostic groups, although designed to show that different varieties of patients may have the same diagnosis, should not be considered as application of *Q* methodology because the correlations between persons were based on standard rating scales designed to show individual differences. There are many such obverse factor analyses, and although they are commonly called *Q* studies, they do not meet Stephenson's criteria. The Bendig and Hamlin (1955) investigation of Rorschach scoring categories is another study of this type.

Perhaps the most valuable applications of factor analysis in *Q* methodology may come from studies of therapeutic phenomena. The possibilities of such an approach were anticipated as early as 1951, when Fiedler published a factor analytic study of differences between therapists from different schools and with different levels of training. Despite the potential of such studies for helping to place psychotherapy on a rational, empirically verifiable basis, only a few students of psychotherapy appear to be ready to study therapeutic phenomena with the systematic planfulness which *Q* methodology could facilitate. In one such study, Nunnally (1955b) had a therapist describe a patient by means of *Q* sorts on eight successive occasions. The factor analysis of these intercorrelations yielded two factors—one concerning relationships with the therapist and the other relating to intrapersonal confidence.

The Peterson, Snyder, Guthrie, and Ray (1958) investigation of

therapeutic biases provides a promising exploration. They approached their study in a sound manner by thoughtfully structuring the sample of statements which comprised their *Q* sorts, systematically using variations of such hypothetical dimensions as direction of gain, attitudes, mode of change, and area of conflict. The sample of therapists who were intercorrelated was drawn from graduates of their own program so that one is not left up in the air with respect to the population of persons to whom the results may be generalized. As is usual in such studies, the factors were interpreted on the basis of the items which received a characteristic sort by persons who had high loadings on the factor. The practice of interpreting persons in terms of item smacks of *R* methodology and reminds us that people are usually more distinguished by their behavior than behavior is distinguished by the people who perform it.

Thrush published an interesting study in 1957. Using a sample of 60 statements descriptive of problems encountered by a counseling agency, the staff made sorts of the level and kind of service each problem would require. These sorts were made in 1952 and again in 1956. On the basis of these sorts, the members of the staff were intercorrelated for each of the years and the two sets of intercorrelations were factor analyzed separately. A comparison of the results indicated that the emphasis in the agency had shifted from vocational counseling to personal adjustment counseling. Although studies of this kind are illuminating, they remind us that we have no rigorous basis for comparing the results of factor analyses to test an exact statistical hypothesis. The question of how one should generalize from a



*Q*-type study is usually disregarded. Conger, Sawrey, and Krause (1956) point to an aspect of this problem in their study of Beck's "The Six Schizophrenias" (1954).

In commenting upon factor analysis in *Q* methodology, one should remember that Stephenson indicated that the correlations should be in part expressive of the *effect* of different kinds of operations. He intended that the intercorrelated variates should, in some manner or another, be regarded as dependent variables in an experimental sense and not merely descriptive dimensions of a static situation. From the standpoint of this emphasis, the Sweetland and Frank (1955) study of ideal psychological adjustment is not a good example of *Q* methodology because its purpose appears to be to describe kinds of psychological adjustment rather than to reveal the effects of certain operations, i.e., it is not a dependency-type analysis. This descriptive use of the *Q*-type factor analysis is not unique to Sweetland and Frank, however; other examples would include Broen's (1957) factor analytic study of religious attitudes.

Many of the samples of statements which have been sorted in *Q* methodology appear to have been somewhat informally assembled, and as a consequence, the analyses performed on the sorts provided by various persons or by the same person under various instructions have an uncertain meaning. We do not know from what parent population of behavior they might conceivably be drawn or from what specific theory they could have been generated. It is probably for this reason that we find relatively few studies where the *Q* sort arrays for an individual or a group of individuals are submitted to an analysis

of variance. This is unfortunate because the difference or similarity between the *Q* sorts of two individuals or the *Q* sorts of an individual under two or more conditions must be explained in terms of the sorted items which comprise the *Q* arrays. If the items had been included in the sample as a priori representatives of theoretically relevant classes of behavior, then the order given to the items could in the case of any given *Q* sort be entered into an analysis of variance. In this way the relative status which the sorter assigned to various a priori classes of items could be revealed. In many studies, however, a defensible a priori classification of behavior with respect to kinds and levels is not possible because the area of inquiry is not well known, no systematic theory can be confidently applied, and in a sense the investigation is exploratory. If, in the study of such an area of behavior, *Q* methodology were indicated, it would seem desirable first to intercorrelate and factor analyze the items in the *R* tradition. Then a sample of statements for *Q* sorts could be arranged so that the various factors could be represented in a balanced design. From such structured samples of statements, the *Q* methodology could be applied in the recommended manner by first factor analyzing the variates (e.g., people) and then explaining the *Q* factors in terms of an analysis of variance of the sorts provided by the variates. The reviewer saw no studies where the domain of behavior was first explored by an *R*-type analysis as a basis for building a structured sample of statements for the *Q* sorts. Where analysis of variance had been applied to *Q* sort arrays, the investigator had carefully structured his sample on an



a priori basis. Such studies are few and tend to be found in the recent literature.

One of the earlier studies involving an analysis of variance was provided in 1956 by Kerlinger who constructed a set of *Q* statements which represent two kinds of educational attitudes interpreted at four different levels each. The levels for each class were then systematically replicated with 10 statements each, so that there were 80 statements in all.

In 1958, Rawn published a study of transference and resistance in psychotherapy. The statements to be sorted conformed with the requirements of a balanced block design involving two levels of resistance and three classes of transference. These categories of class and level could be combined to form six kinds of statements. Each of these types of statements was interpreted in 15 different ways to form the replications, and accordingly the structured sample comprised 90 statements in all. These statements were sorted by different raters and for different sessions of recorded psychotherapy. Because of the way the sample was structured the investigator could perform an analysis of variance for the various sorts as well as factor analyze the intercorrelations among the sorts. His purposes required the analysis of variance only, however.

Perhaps some of the most substantial values to accrue from the point of view known as *Q* methodology may lie in the fact that more of us have become increasingly thoughtful about many matters which we had formerly disregarded or postponed. Possibly one of these neglected matters is the hiatus between the clinicians who continue to be interested in intra-individual differ-

ences and the psychometricians who, acting from the standpoint of interindividual concepts of reliability, have dismissed intra-individual differences as trivial or of no possible consequence.

There is a general tendency for investigators to compute correlation coefficients without giving much thought to the meaning or the determiners of the relationship. *Q* methodology is leading us to think more realistically about features which contribute to the degree of correlation between either subjects or items. If, for example, the sample of items is not homogeneous, it would seem possible for several pairs of persons to be equally correlated with each other but for the various pairs to have their respective correlations because of different items of behavior. As a consequence, none of the items may characterize all of the intercorrelated persons. Presumably a similar kind of situation could exist if items were intercorrelated for a group of persons representing subsamples of different populations. In such a case the correlations found between any two items might vary considerably if they were separately calculated for the various subsamples instead of being calculated for the heterogeneous group. Obviously, the investigator is on shaky ground when he assumes that a correlation based on one sample is descriptive of some other sample which is comprised in some different manner. The composition of the sample with respect to persons can obviously influence the correlation between items, or the composition of a sample with respect to items could influence the correlation between persons.

Some aspects of this problem of subject homogeneity were discussed

by Block in 1955, and in the same year Nunnally (1955a) described an hypothetical matrix where sample heterogeneity with respect to persons resulted in very low correlations between variables while the obverse type correlations between the individuals were very high. Nunnally implied that ipsative scores are particularly valuable in yielding *Q*-type correlations which could reveal trends not apparent from *R*-type analyses. The way in which this matter depends upon the homogeneity of samples and the way in which it may be related to type of scale were not made explicit, however.

The growing interest in *Q* procedure has generated several methodological studies. Cohen (1957) has prepared a monograph which permits the investigator to read correlation coefficients between *Q* sorts, and Creaser (1955) has recommended a way for determining the amount an item should be weighted with respect to a given factor.

Goodling and Guthrie (1956) point out that the sample of items for *Q* sort should be selected in such a way as to provide maximum intersubject variability and minimum intrasubject variability. The question of intrasubject variability is one aspect of the reliability question, and this has been attacked directly by some investigators. For example, Hilden (1958) describes a sampling experiment where he begins with a universe of 1,575 statements from which he has randomly drawn 20 samples of 50 statements each. He had four graduate students provide self-ideal sorts for each of the 20 random sets and for the total population as well. The various scores, e.g., self-ideal, from any one set were correlated with each other, and the respective correlations were determined for the popu-

lation. When the correlations for the random sets were compared with the correlations for the parent population, no reliable differences were found. From this one might infer that when using items such as these, a sample of 50 statements may be sufficient for *Q*-sort purposes.

There appears to be a general tendency among investigators to require their subjects to distribute their *Q* sorts in a quasi-normal fashion. This is in spite of the fact that Stephenson had recommended a flattened, bell-shaped distribution and that subsequent investigators had questioned the desirability of quasi-normal distributions. Jones (1956), for example, had noted that the free sorts of various groups differed appreciably from each other and that no group selected a bell-shaped distribution. Livson and Nichols (1956) had examined this problem from the standpoint of the number of discriminations that various shaped distributions involve, and noted that the more discriminations required, the greater the test-retest reliability of the sort. On the basis of this finding, these authors recommend that the *Q*-sort distribution should be rectangular. The issue of forced vs. unforced sorts has been discussed in numerous contexts, and no final agreement seems to have been reached. For example, Jones points out that there is no one preferred distribution, and Block (1956) believes, on the basis of his comparisons, that the forced sort method is equal or superior to free sorts.

Whether *Q* methodology will, as Stephenson proposed, create a psychology of the individual remains to be seen. From the standpoint of psychometry with its emphasis on individual differences or from the standpoint of psychoanalysis with its

avoidance of formal instrumentation, Q methodology and the devices it includes do not provide an orthodox approach to the study of the individual. Certainly those particular psychologists who profess to be interested primarily in the individual have not rushed to apply this method to material which is still handled on an anecdotal or case history basis. Nevertheless, Q method's primary contributions to psychology appear

to be in the study of psychotherapy and the related study of persons with personality disorders, and there are indications that this methodological emphasis can contribute to a broad study of personality and numerous related social problems. The growing acceptance of this methodological emphasis again reminds us that psychologists require flexible methods for their researches and will not wait for any orthodoxy.

## REFERENCES

- BECK, S. J. The six schizophrenias. *Res. Monogr. Amer. Orthopsychiat. Ass.*, 1954, No. 6.
- BENDIG, A. W., & HAMLIN, R. M. The psychiatric validity of an inverted factor analysis of Rorschach scoring categories. *J. consult. Psychol.*, 1955, 19, 183-188.
- BLOCK, J. The difference between Q and R. *Psychol. Rev.*, 1955, 62, 356-358.
- BLOCK, J. A comparison of the forced and unforced Q-sorting procedures. *Educ. psychol. Measmt.*, 1956, 16, 481-493.
- BLOCK, J. A comparison between ipsative and normative ratings of personality. *J. abnorm. soc. Psychol.*, 1957, 54, 50-54.
- BLOCK, J., & PETERSON, P. Some personality correlates of confidence, caution, and speed in a decision situation. *J. abnorm. soc. Psychol.*, 1955, 51, 34-41.
- BROEN, W. E., JR. A factor-analytic study of religious attitudes. *J. abnorm. soc. Psychol.*, 1957, 54, 176-179.
- BUTLER, J. M., & HAIGH, G. V. Changes in the relation between self-concepts and ideal concepts. In C. R. Rogers & Rosalind F. Dymond (Eds.), *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954. Pp. 55-75.
- CAPLAN, S. W. The effect of group counseling on junior high school boys' concepts of themselves in school. *J. counsel. Psychol.*, 1957, 4, 124-128.
- CARTWRIGHT, ROSALIND D. Effects of psychotherapy on self-consistency. *J. counsel. Psychol.*, 1957, 4, 15-22.
- CATTELL, R. B. Psychological measurement: Ipsative, normative, and interactive. *Psychol. Rev.*, 1944, 51, 292-303.
- CHASE, P. H. Self concepts in adjusted and maladjusted hospital patients. *J. consult. Psychol.*, 1957, 21, 495-497.
- COHEN, J. An aid in the computation of correlations based on Q sorts. *Psychol. Bull.*, 1957, 54, 138-139.
- CONGER, J. J., SAWREY, W. L., & KRAUSE, L. F. A reanalysis of Beck's "six schizophrenias." *J. consult. Psychol.*, 1956, 20, 83-87.
- CORSINI, R. J. Understanding and similarity in marriage. *J. abnorm. soc. Psychol.*, 1956, 52, 327-332.
- CREASER, J. W. An aid in calculating Q-sort factor-arrays. *J. clin. Psychol.*, 1955, 11, 195-196.
- DOLEYS, E. J., & KREGARMAN, J. Construct validity of the Chicago Q-sort: Frustration tolerance. *J. clin. Psychol.*, 1959, 15, 177-179.
- EDWARDS, A. L. Social desirability and Q sorts. *J. consult. Psychol.*, 1955, 19, 462.
- ENGEL, MARY. The stability of the self-concept in adolescence. *J. abnorm. soc. Psychol.*, 1959, 58, 211-215.
- EPSTEIN, S., & SMITH, R. Repression and insight as related to reaction to cartoons. *J. consult. Psychol.*, 1956, 20, 391-395.
- FAGAN, J., & GUTHRIE, G. M. Perception of self and of normality in schizophrenics. *J. clin. Psychol.*, 1959, 15, 203-207.
- FIEDLER, F. E. Factor analyses of psychoanalytic, nondirective, and Adlerian therapeutic relationships. *J. consult. Psychol.*, 1951, 15, 32-38.
- FISK, D. W., & VAN BUSKIRK, C. The stability of interpretations of sentence completion tests. *J. consult. Psychol.*, 1959, 23, 177-180.
- FRIEDMAN, I. Phenomenal, ideal, and projected conceptions of self. *J. abnorm. soc. Psychol.*, 1955, 51, 611-615.
- GOODLING, R. A., & GUTHRIE, G. M. Some practical considerations in Q-sort item selection. *J. counsel. Psychol.*, 1956, 3, 70-72.
- GUILFORD, J. P. *Psychometric methods*. New York: McGraw-Hill, 1954.

- HELFAND, I. Role taking in schizophrenia. *J. consult. Psychol.*, 1956, 20, 37-41.
- HILDEN, A. H. Q-sort correlation: Stability and random choice of statements. *J. consult. Psychol.*, 1958, 22, 45-50.
- JONES, A. Distributions of traits in current Q-sort methodology. *J. abnorm. soc. Psychol.*, 1956, 53, 90-95.
- KALIS, BETTY L., & BENNETT, LILLIAN F. The assessment of communication: The relation of clinical improvement to measured changes in communicative behavior. *J. consult. Psychol.*, 1957, 21, 10-14.
- KERLINGER, F. N. The attitude structure of the individual: A Q-study of the educational attitudes of professors and laymen. *Genet. psychol. Monogr.*, 1956, 53, 283-329.
- KOGAN, W. S., QUINN, R., AX, A. F., & RIPLEY, H. S. Some methodological problems in the quantification of clinical assessment by Q array. *J. consult. Psychol.*, 1957, 21, 57-62.
- LEVY, L. H. The meaning and generality of perceived actual-ideal discrepancies. *J. consult. Psychol.*, 1956, 20, 396-398.
- LIVSON, N. H., & NICHOLS, T. F. Discrimination and reliability in Q-sort personality descriptions. *J. abnorm. soc. Psychol.*, 1956, 52, 159-165.
- MORSH, J. E. The Q-sort technique as a group measure. *Educ. psychol. Measmt.*, 1955, 15, 390-395.
- NAHINSKY, I. D. The relationship between the self-concept and the ideal-self concept as a measure of adjustment. *J. clin. Psychol.*, 1958, 14, 360-364.
- NUNNALLY, J. C. Some uses for "transpose" factor design in assessment research. *Educ. psychol. Measmt.*, 1955, 15, 240-245. (a)
- NUNNALLY, J. C. A systematic approach to the construction of hypotheses about the process of psychotherapy. *J. consult. Psychol.*, 1955, 19, 17-20. (b)
- PETERSON, A. O. D., SNYDER, W. U., GUTHRIE, G. M., & RAY, W. S. Therapist factors: An exploratory investigation of therapeutic biases. *J. counsel. Psychol.*, 1958, 5, 169-173.
- RAWN, M. L. An experimental study of transference and resistance phenomena in psychoanalytically oriented psychotherapy. *J. clin. Psychol.*, 1958, 14, 418-425.
- REVIE, V. A. The effect of psychological case work on the teacher's concept of the pupil. *J. counsel. Psychol.*, 1956, 3, 125-129.
- REZNIKOFF, M., & TOOMEY, LAURA C. The weighted Q sort: A procedure for quantitatively estimating emotional disturbance and personality change. *J. consult. Psychol.*, 1958, 22, 187-190.
- ROGERS, A. H. The self-concept in paranoid schizophrenia. *J. clin. Psychol.*, 1958, 14, 365-366.
- ROGERS, C. R., & DYMOND, ROSALIND F. *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954.
- SHEPHERD, IRMA L., & GUTHRIE, G. M. Attitudes of mothers of schizophrenic patients. *J. clin. Psychol.*, 1959, 15, 212-215.
- SHONTZ, F. C. Evaluative conceptualizations as the basis for clinical judgments. *J. consult. Psychol.*, 1956, 20, 223-226.
- STEPHENSON, W. *The study of behavior*. Chicago: Univer. Chicago Press, 1953.
- STEWART, L. H. Manifest anxiety and mother-son identification. *J. clin. Psychol.*, 1958, 14, 382-384.
- SWEETLAND, A., & FRANK, G. A study of ideal psychological adjustment. *J. clin. Psychol.*, 1955, 11, 391-394.
- TAYLOR, D. M. Changes in the self concept without psychotherapy. *J. consult. Psychol.*, 1955, 19, 205-209.
- THRUSH, R. S. An agency in transition: The case study of a counseling center. *J. counsel. Psychol.*, 1957, 4, 183-190.
- TURNER, R. H., & VANDERLIPPE, R. H. Self-ideal congruence as an index of adjustment. *J. abnorm. soc. Psychol.*, 1958, 57, 202-206.
- WHITING, J. F. Needs, values, perceptions, and the nurse-patient relationship. *J. clin. Psychol.*, 1959, 15, 146-150.

(Received January 4, 1960)

## PSYCHOTHERAPY AS A LEARNING PROCESS

ALBERT BANDURA

*Stanford University*

While it is customary to conceptualize psychotherapy as a learning process, few therapists accept the full implications of this position. Indeed, this is best illustrated by the writings of the learning theorists themselves. Most of our current methods of psychotherapy represent an accumulation of more or less uncontrolled clinical experiences and, in many instances, those who have written about psychotherapy in terms of learning theory have merely substituted a new language; the practice remains essentially unchanged (Dollard, Auld, & White, 1954; Dollard & Miller, 1950; Shoben, 1949).

If one seriously subscribes to the view that psychotherapy is a learning process, the methods of treatment should be derived from our knowledge of learning and motivation. Such an orientation is likely to yield new techniques of treatment which, in many respects, may differ markedly from the procedures currently in use.

Psychotherapy rests on a very simple but fundamental assumption, i.e., human behavior is modifiable through psychological procedures. When skeptics raise the question, "Does psychotherapy work?" they may be responding in part to the mysticism that has come to surround the term. Perhaps the more meaningful question, and one which avoids the surplus meanings associated with the term "psychotherapy," is as follows: Can human behavior be modified through psychological means and if so, what are the learning mechanisms that mediate behavior change?

In the sections that follow, some of these learning mechanisms will be discussed, and studies in which systematic attempts have been made to apply these principles of learning to the area of psychotherapy will be reviewed. Since learning theory itself is still somewhat incomplete, the list of psychological processes by which changes in behavior can occur should not be regarded as exhaustive, nor are they necessarily without overlap.

### COUNTERCONDITIONING

Of the various treatment methods derived from learning theory, those based on the principle of counterconditioning have been elaborated in greatest detail. Wolpe (1954, 1958, 1959) gives a thorough account of this method, and additional examples of cases treated in this manner are provided by Jones (1956), Lazarus and Rachman (1957), Meyer (1957), and Rachman (1959). Briefly, the principle involved is as follows: if strong responses which are incompatible with anxiety reactions can be made to occur in the presence of anxiety evoking cues, the incompatible responses will become attached to these cues and thereby weaken or eliminate the anxiety responses.

The first systematic psychotherapeutic application of this method was reported by Jones (1924b) in the treatment of Peter, a boy who showed severe phobic reactions to animals, fur objects, cotton, hair, and mechanical toys. Counterconditioning was achieved by feeding the child in the presence of initially small but gradually increasing anxiety-arousing



stimuli. A rabbit in a cage was placed in the room at some distance so as not to disturb the boy's eating. Each day the rabbit was brought nearer to the table and eventually removed from the cage. During the final stage of treatment, the rabbit was placed on the feeding table and even in Peter's lap. Tests of generalization revealed that the fear responses had been effectively eliminated, not only toward the rabbit, but toward the previously feared furry objects as well.

In this connection, it would be interesting to speculate on the diagnosis and treatment Peter would have received had he been seen by Melanie Klein (1949) rather than by Mary Cover Jones!

It is interesting to note that while both Shoben (1949) and Wolpe (1958) propose a therapy based on the principle of counterconditioning, their treatment methods are radically different. According to Shoben, the patient discusses and thinks about stimulus situations that are anxiety provoking in the context of an interpersonal situation which simultaneously elicits positive affective responses from the patient. The therapeutic process consists in connecting the anxiety provoking stimuli, which are symbolically reproduced, with the comfort reaction made to the therapeutic relationship.

Shoben's paper represents primarily a counterconditioning interpretation of the behavior changes brought about through conventional forms of psychotherapy since, apart from highlighting the role of positive emotional reactions in the treatment process, no new techniques deliberately designed to facilitate relearning through counterconditioning are proposed.

This is not the case with Wolpe, who has made a radical departure

from tradition. In his treatment, which he calls reciprocal inhibition, Wolpe makes systematic use of three types of responses which are antagonistic to, and therefore inhibitory of, anxiety. These are: assertive or approach responses, sexual responses, and relaxation responses.

On the basis of historical information, interview data, and psychological test responses, the therapist constructs an anxiety hierarchy, a ranked list of stimuli to which the patient reacts with anxiety. In the case of desensitization based on relaxation, the patient is hypnotized and given relaxation suggestions. He is then asked to imagine a scene representing the weakest item on the anxiety hierarchy and, if the relaxation is unimpaired, this is followed by having the patient imagine the next item on the list, and so on. Thus, the anxiety cues are gradually increased from session to session until the last phobic stimulus can be presented without impairing the relaxed state. Through this procedure, relaxation responses eventually come to be attached to the anxiety evoking stimuli.

Wolpe reports remarkable therapeutic success with a wide range of neurotic reactions treated on this counterconditioning principle. He also contends that the favorable outcomes achieved by the more conventional psychotherapeutic methods may result from the reciprocal inhibition of anxiety by strong positive responses evoked in the patient-therapist relationship.

Although the counterconditioning method has been employed most extensively in eliminating anxiety-motivated avoidance reactions and inhibitions, it has been used with some success in reducing maladaptive approach responses as well. In the



latter case, the goal object is repeatedly associated with some form of aversive stimulus.

Raymond (1956), for example, used nausea as the aversion experience in the treatment of a patient who presented a fetish for handbags and perambulators which brought him into frequent contact with the law in that he repeatedly smeared mucus on ladies' handbags and destroyed perambulators by running into them with his motorcycle. Though the patient had undergone psychoanalytic treatment, and was fully aware of the origin and the sexual significance of his behavior, nevertheless, the fetish persisted.

The treatment consisted of showing the patient a collection of handbags, perambulators, and colored illustrations just before the onset of nausea produced by injections of apomorphine. The conditioning was repeated every 2 hours day and night for 1 week plus additional sessions 8 days and 6 months later.

Raymond reports that, not only was the fetish successfully eliminated, but also the patient showed a vast improvement in his social (and legal) relationships, was promoted to a more responsible position in his work, and no longer required the fetish fantasies to enable him to have sexual intercourse.

Nauseant drugs, especially emetine, have also been utilized as the unconditioned stimulus in the aversion treatment of alcoholism (Thirrmann, 1949; Thompson & Bielinski, 1953; Voegtlen, 1940; Wallace, 1949). Usually 8 to 10 treatments in which the sight, smell, and taste of alcohol is associated with the onset of nausea is sufficient to produce abstinence. Of 1,000 or more cases on whom adequate follow-up data are reported, approximately 60% of the patients

have been totally abstinent following the treatment. Voegtlen (1940) suggests that a few preventive treatments given at an interval of about 6 months may further improve the results yielded by this method.

Despite these encouraging findings, most psychotherapists are unlikely to be impressed since, in their opinion, the underlying causes for the alcoholism have in no way been modified by the conditioning procedure and, if anything, the mere removal of the alcoholism would tend to produce symptom substitution or other adverse effects. A full discussion of this issue will be presented later. In this particular context, however, several aspects of the Thompson and Bielinski (1953) data are worth noting. Among the alcoholic patients whom they treated, six "suffered from mental disorders not due to alcohol or associated deficiency states." It was planned, by the authors, to follow up the aversion treatment with psychotherapy for the underlying psychosis. This, however, proved unnecessary since all but one of the patients, a case of chronic mental deterioration, showed marked improvement and were in a state of remission.

Max (1935) employed a strong electric shock as the aversive stimulus in treating a patient who tended to display homosexual behavior following exposure to a fetishistic stimulus. Both the fetish and the homosexual behavior were removed through a series of avoidance conditioning sessions in which the patient was administered shock in the presence of the fetishistic object.

Wolpe (1958) has also reported favorable results with a similar procedure in the treatment of obsessions.

A further variation of the counter-conditioning procedure has been developed by Mowrer and Mowrer

(1938) for use with enuretic patients. The device consists of a wired bed pad which sets off a loud buzzer and awakens the child as soon as micturition begins. Bladder tension thus becomes a cue for waking up which, in turn, is followed by sphincter contraction. Once bladder pressure becomes a stimulus for the more remote sphincter control response, the child is able to remain dry for relatively long periods of time without wakening.

Mowrer and Mowrer (1938) report complete success with 30 children treated by this method; similarly, Davidson and Douglass (1950) achieved highly successful results with 20 chronic enuretic children (15 cured, 5 markedly improved); of 5 cases treated by Morgan and Witmer (1939), 4 of the children not only gained full sphincter control, but also made a significant improvement in their social behavior. The one child with whom the conditioning approach had failed was later found to have bladder difficulties which required medical attention.

Some additional evidence for the efficacy of this method is provided by Martin and Kubly (1955) who obtained follow-up information from 118 of 220 parents who had treated their children at home with this type of conditioning apparatus. In 74% of the cases, according to the parents' replies, the treatment was successful.

#### EXTINCTION

"When a learned response is repeated without reinforcement the strength of the tendency to perform that response undergoes a progressive decrease" (Dollard & Miller, 1950). Extinction involves the development of inhibitory potential which is composed of two components. The evocation of any reaction generates reactive inhibition ( $I_r$ ) which presumably

dissipates with time. When reactive inhibition (fatigue, etc.) reaches a high point, the cessation of activity alleviates this negative motivational state and any stimuli associated with the cessation of the response become conditioned inhibitors ( $I_i$ ).

One factor that has been shown to influence the rate of extinction of maladaptive and anxiety-motivated behavior is the interval between extinction trials. In general, there tends to be little diminution in the strength of fear-motivated behavior when extinction trials are widely distributed, whereas under massed trials, reactive inhibition builds up rapidly and consequently extinction is accelerated (Calvin, Clifford, Clifford, Bolden, & Harvey, 1956; Edmonson & Amsel, 1954).

An illustration of the application of this principle is provided by Yates (1958) in the treatment of tics. Yates demonstrated, in line with the findings from laboratory studies of extinction under massed and distributed practice, that massed sessions in which the patient performed tics voluntarily followed by prolonged rest to allow for the dissipation of reactive inhibition was the most effective procedure for extinguishing the tics.

It should be noted that the extinction procedure employed by Yates is very similar to Dunlap's method of negative practice, in which the subject reproduces the negative behaviors voluntarily without reinforcement (Dunlap, 1932; Lehner, 1954). This method has been applied most frequently, with varying degrees of success, to the treatment of speech disorders (Fishman, 1937; Meissner, 1946; Rutherford, 1940; Sheehan, 1951; Sheehan & Voas, 1957). If the effectiveness of this psychotherapeutic technique is due primarily to extinction, as suggested by Yates'

study, the usual practice of terminating a treatment session before the subject becomes fatigued (Lehner, 1954), would have the effect of reducing the rate of extinction, and may in part account for the divergent results yielded by this method.

Additional examples of the therapeutic application of extinction procedures are provided by Jones (1955), and most recently by C. D. Williams (1959).

Most of the conventional forms of psychotherapy rely heavily on extinction effects although the therapist may not label these as such. For example, many therapists consider *permissiveness* to be a necessary condition of therapeutic change (Alexander, 1956; Dollard & Miller, 1950; Rogers, 1951). It is expected that when a patient expresses thoughts or feelings that provoke anxiety or guilt and the therapist does not disapprove, criticize, or withdraw interest, the fear or guilt will be gradually weakened or extinguished. The extinction effects are believed to generalize to thoughts concerning related topics that were originally inhibited, and to verbal and physical forms of behavior as well (Dollard & Miller, 1950).

Some evidence for the relationship between permissiveness and the extinction of anxiety is provided in two studies recently reported by Dittes (1957a, 1957b). In one study (1957b) involving an analysis of patient-therapist interaction sequences, Dittes found that permissive responses on the part of the therapist were followed by a corresponding decrease in the patient's anxiety (as measured by the GSR) and the occurrence of avoidance behaviors. A sequential analysis of the therapeutic sessions (Dittes, 1957a), revealed that, at the onset of treatment, sex expressions were accompanied by strong anxiety

reactions; under the cumulative effects of permissiveness, the anxiety gradually extinguished.

In contrast to counterconditioning, extinction is likely to be a less effective and a more time consuming method for eliminating maladaptive behavior (Jones, 1924a; Dollard & Miller, 1950); in the case of conventional interview therapy, the relatively long intervals between interview sessions, and the ritualistic adherence to the 50-minute hour may further reduce the occurrence of extinction effects.

#### DISCRIMINATION LEARNING

Human functioning would be extremely difficult and inefficient if a person had to learn appropriate behavior for every specific situation he encountered. Fortunately, patterns of behavior learned in one situation will transfer or generalize to other similar situations. On the other hand, if a person overgeneralizes from one situation to another, or if the generalization is based on superficial or irrelevant cues, behavior becomes inappropriate and maladaptive.

In most theories of psychotherapy, therefore, discrimination learning, believed to be accomplished through the gaining of awareness or insight, receives emphasis (Dollard & Miller, 1950; Fenichel, 1941; Rogers, 1951; Sullivan, 1953). It is generally assumed that if a patient is aware of the cues producing his behavior, of the responses he is making, and of the reasons that he responds the way he does, his behavior will become more susceptible to verbally-mediated control. Voluntarily guided, discriminative behavior will replace the automatic, overgeneralized reactions.

While this view is widely accepted, as evidenced in the almost exclusive reliance on interview procedures and on interpretative or labeling tech-

niques, a few therapists (Alexander & French, 1946) have questioned the importance attached to awareness in producing modifications in behavior. Whereas most psychoanalysts (Fenichel, 1941), as well as therapists representing other points of view (Fromm-Reichmann, 1950; Sullivan, 1953) consider insight a precondition of behavior change, Alexander and French consider insight or awareness a result of change rather than its cause. That is, as the patient's anxieties are gradually reduced through the permissive conditions of treatment, formerly inhibited thoughts are gradually restored to awareness.

Evidence obtained through controlled laboratory studies concerning the value of awareness in increasing the precision of discrimination has so far been largely negative or at least equivocal (Adams, 1957; Erikson, 1958; Razran, 1949). A study by Lacy and Smith (1954), in which they found aware subjects generalized anxiety reactions less extensively than did subjects who were unaware of the conditioned stimulus provides evidence that awareness may aid discrimination. However, other aspects of their findings (e.g., the magnitude of the anxiety reactions to the generalization stimuli were greater than they were to the conditioned stimulus itself) indicate the need for replication.

If future research continues to demonstrate that awareness exerts little influence on the acquisition, generalization, and modification of behavior, such negative results would cast serious doubt on the value of currently popular psychotherapeutic procedures whose primary aim is the development of insight.

#### METHODS OF REWARD

Most theories of psychotherapy are based on the assumption that the

patient has a repertoire of previously learned positive habits available to him, but that these adaptive patterns are inhibited or blocked by competing responses motivated by anxiety or guilt. The goal of therapy, then, is to reduce the severity of the internal inhibitory controls, thus allowing the healthy patterns of behavior to emerge. Hence, the role of the therapist is to create permissive conditions under which the patient's "normal growth potentialities" are set free (Rogers, 1951). The fact that most of our theories of personality and therapeutic procedures have been developed primarily through work with oversocialized, neurotic patients may account in part for the prevalence of this view.

There is a large class of disorders (the undersocialized, antisocial personalities whose behavior reflects a failure of the socialization process) for whom this model of personality and accompanying techniques of treatment are quite inappropriate (Bandura & Walters, 1959; Schmidberg, 1959). Such antisocial personalities are likely to present *learning deficits*, consequently the goal of therapy is the acquisition of secondary motives and the development of internal restraint habits. That antisocial patients prove unresponsive to psychotherapeutic methods developed for the treatment of oversocialized neurotics has been demonstrated in a number of studies comparing patients who remain in treatment with those who terminate treatment prematurely (Rubenstein & Lorr, 1956). It is for this class of patients that the greatest departures from traditional treatment methods is needed.

While counterconditioning, extinction, and discrimination learning may be effective ways of removing neurotic inhibitions, these methods may be of relatively little value in develop-

ing new positive habits. Primary and secondary rewards in the form of the therapist's interest and approval may play an important, if not indispensable, role in the treatment process. Once the patient has learned to want the interest and approval of the therapist, these rewards may then be used to promote the acquisition of new patterns of behavior. For certain classes of patients such as schizophrenics (Atkinson, 1957; Peters, 1953; Robinson, 1957) and delinquents (Cairns, 1959), who are either unresponsive to, or fearful of, social rewards, the therapist may have to rely initially on primary rewards in the treatment process.

An ingenious study by Peters and Jenkins (1954) illustrates the application of this principle in the treatment of schizophrenic patients. Chronic patients from closed wards were administered subshock injections of insulin designed to induce the hunger drive. The patients were then encouraged to solve a series of graded problem tasks with fudge as the reward. This program was followed 5 days a week for 3 months.

Initially the tasks involved simple mazes and obstruction problems in which the patients obtained the food reward directly upon successful completion of the problem. Tasks of gradually increasing difficulty were then administered involving multiple-choice learning and verbal-reasoning problems in which the experimenter personally mediated the primary rewards. After several weeks of such problem solving activities the insulin injections were discontinued and social rewards, which by this time had become more effective, were used in solving interpersonal problems that the patients were likely to encounter in their daily activities both inside and outside the hospital setting.

Comparison of the treated group with control groups, designed to isolate the effects of insulin and special attention, revealed that the patients in the reward group improved significantly in their social relationships in the hospital, whereas the patients in the control groups showed no such change.

King and Armitage (1958) report a somewhat similar study in which severely withdrawn schizophrenic patients were treated with operant conditioning methods; candy and cigarettes served as the primary rewards for eliciting and maintaining increasingly complex forms of behavior, i.e., psychomotor, verbal, and interpersonal responses. Unlike the Peters and Jenkins study, no attempt was made to manipulate the level of primary motivation.

An interesting feature of the experimental design was the inclusion of a group of patients who were treated with conventional interview therapy, as well as a recreational therapy and a no-therapy control group. It was found that the operant group, in relation to similar patients in the three control groups, made significantly more clinical improvement.

Skinner (1956b) and Lindsley (1956) working with adult psychotics, and Ferster (1959) working with autistic children, have been successful in developing substantial amounts of reality-oriented behavior in their patients through the use of reward. So far their work has been concerned primarily with the effect of schedules of reinforcement on the rate of evocation of simple impersonal reactions. There is every indication, however, that by varying the contingency of the reward (e.g., the patient must respond in certain specified ways to the behavior of another individual in order to produce the reward) adap-



tive interpersonal behaviors can be developed as well (Azran & Lindsley, 1956).

The effectiveness of social reinforcers in modifying behavior has been demonstrated repeatedly in verbal conditioning experiments (Krasner, 1958; Salzinger, 1959). Encouraged by these findings, several therapists have begun to experiment with operant conditioning as a method of treatment in its own right (Tilton, 1956; Ullman, Krasner, & Collins, *in press*; R. I. Williams, 1959); the operant conditioning studies cited earlier are also illustrative of this trend.

So far the study of generalization and permanence of behavior changes brought about through operant conditioning methods has received relatively little attention and the scanty data available are equivocal (Rogers, 1960; Sarason, 1957; Weide, 1959). The lack of consistency in results is hardly surprising considering that the experimental manipulations in many of the conditioning studies are barely sufficient to demonstrate conditioning effects, let alone generalization of changes to new situations. On the other hand, investigators who have conducted more intensive reinforcement sessions, in an effort to test the efficacy of operant conditioning methods as a therapeutic technique, have found significant changes in patients' interpersonal behavior in extra-experimental situations (King & Armitage, 1958; Peters & Jenkins, 1954; Ullman *et al.*, *in press*). These findings are particularly noteworthy since the response classes involved are similar to those psychotherapists are primarily concerned in modifying through interview forms of treatment. If the favorable results yielded by these studies are replicated in future investigations, it is likely that the next few years will witness an increas-

ing reliance on conditioning forms of psychotherapy, particularly in the treatment of psychotic patients.

At this point it might also be noted that, consistent with the results from verbal conditioning experiments, content analyses of psychotherapeutic interviews (Bandura, Lipsher, & Miller, 1960; Murray, 1956) suggest that many of the changes observed in psychotherapy, at least insofar as the patients' verbal behavior is concerned, can be accounted for in terms of the therapists' direct, although usually unwitting, reward and punishment of the patients' expressions.

#### PUNISHMENT

While positive habits can be readily developed through reward, the elimination of socially disapproved habits, which becomes very much an issue in the treatment of antisocial personalities, poses a far more complex problem.

The elimination of socially disapproved behaviors can be accomplished in several ways. They may be consistently unrewarded and thus extinguished. However, antisocial behavior, particularly of an extreme form, cannot simply be ignored in the hope that it will gradually extinguish. Furthermore, since the successful execution of antisocial acts may bring substantial material rewards as well as the approval and admiration of associates, it is extremely unlikely that such behavior would ever extinguish.

Although punishment may lead to the rapid disappearance of socially disapproved behavior, its effects are far more complex (Estes, 1944; Solomon, Kamin, & Wynne, 1953). If a person is punished for some socially disapproved habit, the impulse to perform the act becomes, through its association with punishment, a stimulus for anxiety. This anxiety



then motivates competing responses which, if sufficiently strong, prevent the occurrence of, or inhibit, the disapproved behavior. Inhibited responses may not, however, thereby lose their strength, and may reappear in situations where the threat of punishment is weaker. Punishment may, in fact, prevent the extinction of a habit; if a habit is completely inhibited, it cannot occur and therefore cannot go unrewarded.

Several other factors point to the futility of punishment as a means of correcting many antisocial patterns. The threat of punishment is very likely to elicit conformity; indeed, the patient may obligingly do whatever he is told to do in order to avoid immediate difficulties. This does not mean, however, that he has acquired a set of sanctions that will be of service to him once he is outside the treatment situation. In fact, rather than leading to the development of internal controls, such methods are likely only to increase the patient's reliance on external restraints. Moreover, under these conditions, the majority of patients will develop the attitude that they will do only what they are told to do—and then often only half-heartedly—and that they will do as they please once they are free from the therapist's supervision (Bandura & Walters, 1959).

In addition, punishment may serve only to intensify hostility and other negative motivations and thus may further instigate the antisocial person to display the very behaviors that the punishment was intended to bring under control.

Mild aversive stimuli have been utilized, of course, in the treatment of voluntary patients who express a desire to rid themselves of specific debilitating conditions.

Liversedge and Sylvester (1955), for example, successfully treated

seven cases of writer's cramp by means of a retraining procedure involving electric shock. In order to remove tremors, one component of the motor disorder, the patients were required to insert a stylus into a series of progressively smaller holes; each time the stylus made contact with the side of the hole the patients received a mild shock. The removal of the spasm component of the disorder was obtained in two ways. First, the patients traced various line patterns (similar to the movements required in writing) on a metal plate with a stylus, and any deviation from the path produced a shock. Following training on the apparatus, the subjects then wrote with an electrified pen which delivered a shock whenever excessive thumb pressure was applied.

Liversedge and Sylvester report that following the retraining the patients were able to resume work; a follow-up several months later indicated that the improvement was being maintained.

The aversive forms of therapy, described earlier in the section on counterconditioning procedures, also make use of mild punishment.

#### SOCIAL IMITATION

Although a certain amount of learning takes place through direct training and reward, a good deal of a person's behavior repertoire may be acquired through imitation of what he observes in others. If this is the case, social imitation may serve as an effective vehicle for the transmission of prosocial behavior patterns in the treatment of antisocial patients.

Merely providing a model for imitation is not, however, sufficient. Even though the therapist exhibits the kinds of behaviors that he wants the patient to learn, this is likely to have little influence on him if he

rejects the therapist as a model. Affectional nurturance is believed to be an important precondition for imitative learning to occur, in that affectional rewards increase the secondary reinforcing properties of the model, and thus predispose the imitator to pattern his behavior after the rewarding person (Mowrer, 1950; Sears, 1957; Whiting, 1954). Some positive evidence for the influence of social rewards on imitation is provided by Bandura and Huston (in press) in a recent study of identification as a process of incidental imitation.

In this investigation preschool children performed an orienting task but, unlike most incidental learning studies, the experimenter performed the diverting task as well, and the extent to which the subjects patterned their behavior after that of the experimenter-model was measured.

A two-choice discrimination problem similar to the one employed by Miller and Dollard (1941) in their experiments of social imitation was used as the diverting task. On each trial, one of two boxes was loaded with two rewards (small multicolor pictures of animals) and the object of the game was to guess which box contained the stickers. The experimenter-model (*M*) always had her turn first and in each instance chose the reward box. During *M*'s trial, the subject remained at the starting point where he could observe the *M*'s behavior. On each discrimination trial *M* exhibited certain verbal, motor, and aggressive patterns of behavior that were totally irrelevant to the task to which the subject's attention was directed. At the starting point, for example, *M* made a verbal response and then marched slowly toward the box containing the stickers, repeating, "March, march, march." On the lid of each box was a

rubber doll which *M* knocked off aggressively when she reached the designated box. She then paused briefly, remarked, "Open the box," removed one sticker, and pasted it on a pastoral scene which hung on the wall immediately behind the boxes. The subject then took his turn and the number of *M*'s behaviors performed by the subject was recorded.

A control group was included in order to, (a) provide a check on whether the subjects' performances reflected genuine imitative learning or merely the chance occurrence of behaviors high in the subjects' response hierarchies, and (b) to determine whether subjects would adopt certain aspects of *M*'s behavior which involved considerable delay in reward. With the controls, therefore, *M* walked to the box, choosing a highly circuitous route along the sides of the experimental room; instead of aggressing toward the doll, she lifted it gently off the container.

The results of this study indicate that, insofar as preschool children are concerned, a good deal of incidental imitation of the behaviors displayed by an adult model does occur. Of the subjects in the experimental group, 88% adopted the *M*'s aggressive behavior, 44% imitated the marching, and 28% reproduced *M*'s verbalizations. In contrast, none of the control subjects behaved aggressively, marched, or verbalized, while 75% of the controls imitated the circuitous route to the containers.

In order to test the hypothesis that children who experience a rewarding relationship with an adult model adopt more of the model's behavior than do children who experience a relatively distant and cold relationship, half the subjects in the experiment were assigned to a nurturant condition; the other half of the subjects to a nonnurturant condition.

During the nurturant sessions, which preceded the incidental learning, *M* played with subject, she responded readily to the subject's bids for attention, and in other ways fostered a consistently warm and rewarding interaction with the child. In contrast, during the nonnurturant sessions, the subject played alone while *M* busied herself with paperwork at a desk in the far corner of the room.

Consistent with the hypothesis, it was found that subjects who experienced the rewarding interaction with *M* adopted significantly more of *M*'s behavior than did subjects who were in the nonnurturance condition.

A more crucial test of the transmission of behavior patterns through the process of social imitation involves the delayed generalization of imitative responses to new situations in which the model is absent. A study of this type just completed, provides strong evidence that observation of the cues produced by the behavior of others is an effective means of eliciting responses for which the original probability is very low (Bandura, Ross, & Ross, in press).

Empirical studies of the correlates of strong and weak identification with parents, lend additional support to the theory that rewards promote imitative learning. Boys whose fathers are highly rewarding and affectionate have been found to adopt the father-role in doll-play activities (Sears, 1953), to show father-son similarity in response to items on a personality questionnaire (Payne & Mussen, 1956), and to display masculine behaviors (Mussen & Distler, 1956, 1960) to a greater extent than boys whose fathers are relatively cold and nonrewarding.

The treatment of older unsocialized delinquents is a difficult task, since they are relatively self-sufficient and do not readily seek involvement with

a therapist. In many cases, socialization can be accomplished only through residential care and treatment. In the treatment home, the therapist can personally administer many of the primary rewards and mediate between the boys' needs and gratifications. Through the repeated association with rewarding experiences for the boy, many of the therapist's attitudes and actions will acquire secondary reward value, and thus the patient will be motivated to reproduce these attitudes and actions in himself. Once these attitudes and values have been thus accepted, the boy's inhibition of antisocial tendencies will function independently of the therapist.

While treatment through social imitation has been suggested as a method for modifying antisocial patterns, it can be an effective procedure for the treatment of other forms of disorders as well. Jones (1924a), for example, found that the social example of children reacting normally to stimuli feared by another child was effective, in some instances, in eliminating such phobic reactions. In fact, next to counterconditioning, the method of social imitation proved to be most effective in eliminating inappropriate fears.

There is some suggestive evidence that by providing high prestige models and thus increasing the reinforcement value of the imitator's behavior, the effectiveness of this method in promoting favorable adjustment patterns of behavior may be further increased (Jones, 1924a; Mausner, 1953, 1954; Miller & Dollard, 1941).

During the course of conventional psychotherapy, the patient is exposed to many incidental cues involving the therapist's values, attitudes, and patterns of behavior. They are incidental only because they are usually

considered secondary or irrelevant to the task of resolving the patient's problems. Nevertheless, some of the changes observed in the patient's behavior may result, not so much from the intentional interaction between the patient and the therapist, but rather from active learning by the patient of the therapist's attitudes and values which the therapist never directly attempted to transmit. This is partially corroborated by Rosenthal (1955) who found that, in spite of the usual precautions taken by therapists to avoid imposing their values on their clients, the patients who were judged as showing the greatest improvement changed their moral values (in the areas of sex, aggression, and authority) in the direction of the values of their therapists, whereas patients who were unimproved became less like the therapist in values.

#### FACTORS IMPEDING INTEGRATION

In reviewing the literature on psychotherapy, it becomes clearly evident that learning theory and general psychology have exerted a remarkably minor influence on the practice of psychotherapy and, apart from the recent interest in Skinner's operant conditioning methods (Krasner, 1955; Skinner, 1953), most of the recent serious attempts to apply learning principles to clinical practice have been made by European psychotherapists (Jones, 1956; Lazarus & Rachman, 1957; Liversedge & Sylvester, 1955; Meyer, 1957; Rachman, 1959; Raymond, 1956; Wolpe, 1958; Yates, 1958). This isolation of the methods of treatment from our knowledge of learning and motivation will continue to exist for some time since there are several prevalent attitudes that impede adequate integration.

In the first place, the deliberate use of the principles of learning in the

modification of human behavior implies, for most psychotherapists, manipulation and control of the patient, and control is seen by them as antihumanistic and, therefore, bad. Thus, advocates of a learning approach to psychotherapy are often charged with treating human beings as though they were rats or pigeons and of leading on the road to Orwell's 1984.

This does not mean that psychotherapists do not influence and control their patients' behavior. On the contrary. In any interpersonal interaction, and psychotherapy is no exception, people influence and control one another (Frank, 1959; Skinner, 1956a). Although the patient's control of the therapist has not as yet been studied (such control is evident when patients subtly reward the therapist with interesting historical material and thereby avoid the discussion of their current interpersonal problems), there is considerable evidence that the therapist exercises personal control over his patients. A brief examination of interview protocols of patients treated by therapists representing differing theoretical orientations, clearly reveals that the patients have been thoroughly conditioned in their therapists' idiosyncratic languages. Client-centered patients, for example, tend to produce the client-centered terminology, theory, and goals, and their interview content shows little or no overlap with that of patients seen in psychoanalysis who, in turn, tend to speak the language of psychoanalytic theory (Heine, 1950). Even more direct evidence of the therapists' controlling influence is provided in studies of patient-therapist interactions (Bandura et al., 1960; Murray, 1956; Rogers, 1960). The results of these studies show that the therapist not only controls the patient by reward-

ing him with interest and approval when the patient behaves in a fashion the therapist desires, but that he also controls through punishment, in the form of mild disapproval and withdrawal of interest, when the patient behaves in ways that are threatening to the therapist or run counter to his goals.

One difficulty in understanding the changes that occur in the course of psychotherapy is that the independent variable, i.e., the therapist's behavior, is often vaguely or only partially defined. In an effort to minimize or to deny the therapist's directive influence on the patient, the therapist is typically depicted as a "catalyst" who, in some mysterious way, sets free positive adjustive patterns of behavior or similar outcomes usually described in very general and highly socially desirable terms.

It has been suggested, in the material presented in the preceding sections, that many of the changes that occur in psychotherapy derive from the unwitting application of well-known principles of learning. However, the occurrence of the necessary conditions for learning is more by accident than by intent and, perhaps, a more deliberate application of our knowledge of the learning process to psychotherapy would yield far more effective results.

The predominant approach in the development of psychotherapeutic procedures has been the "school" approach. A similar trend is noted in the treatment methods being derived from learning theory. Wolpe, for example, has selected the principle of counterconditioning and built a "school" of psychotherapy around it; Dollard and Miller have focused on extinction and discrimination learning; and the followers of Skinner rely almost entirely on methods of reward. This stress on a few learning

principles at the expense of neglecting other relevant ones will serve only to limit the effectiveness of psychotherapy.

A second factor that may account for the discontinuity between general psychology and psychotherapeutic practice is that the model of personality to which most therapists subscribe is somewhat dissonant with the currently developing principles of behavior.

In their formulations of personality functioning, psychotherapists are inclined to appeal to a variety of inner explanatory processes. In contrast, learning theorists view the organism as a far more mechanistic and simpler system, and consequently their formulations tend to be expressed for the most part in terms of antecedent-consequent relationships without reference to inner states.

Symptoms are learned S-R connections; once they are extinguished or deconditioned treatment is complete. Such treatment is based exclusively on present factors; like Lewin's theory, this one is a-historical. Non-verbal methods are favored over verbal ones, although a minor place is reserved for verbal methods of extinction and reconditioning. Concern is with *function*, not with *content*. The main difference between the two theories arises over the question of "symptomatic" treatment. According to orthodox theory, this is useless unless the underlying complexes are attacked. According to the present theory, there is no evidence for these putative complexes, and symptomatic treatment is all that is required (Eysenck, 1957, pp. 267-268). (Quoted by permission of Frederick A. Praeger, Inc.)

Changes in behavior brought about through such methods as counterconditioning are apt to be viewed by the "dynamically-oriented" therapist, as being not only superficial, "symptomatic" treatment, in that the basic underlying instigators of the behavior remain unchanged, but also potentially dangerous, since the direct elimination of a symptom may



precipitate more seriously disturbed behavior.

This expectation receives little support from the generally favorable outcomes reported in the studies reviewed in this paper. In most cases where follow-up data were available to assess the long-term effects of the therapy, the patients, many of whom had been treated by conventional methods with little benefit, had evidently become considerably more effective in their social, vocational, and psychosexual adjustment. On the whole the evidence, while open to error, suggests that no matter what the origin of the maladaptive behavior may be, a change in behavior brought about through learning procedures may be all that is necessary for the alleviation of most forms of emotional disorders.

As Mowrer (1950) very aptly points out, the "symptom-underlying cause" formulation may represent inappropriate medical analogizing. Whether or not a given behavior will be considered normal or a symptom of an underlying disturbance will depend on whether or not somebody objects to the behavior. For example, aggressiveness on the part of children may be encouraged and considered a sign of healthy development by the parents, while the same behavior is viewed by school authorities and society as a symptom of a personality disorder (Bandura & Walters, 1959). Furthermore, behavior considered to be normal at one stage in development may be re-regarded as a "symptom of a personality disturbance" at a later period. In this connection it is very appropriate to repeat Mowrer's (1950) query: "And when does persisting behavior of this kind suddenly cease to be normal and become a symptom" (p. 474).

Thus, while a high fever is generally considered a sign of an underlying

disease process regardless of when or where it occurs, whether a specific behavior will be viewed as normal or as a symptom of an underlying pathology is not independent of who makes the judgement, the social context in which the behavior occurs, the age of the person, as well as many other factors.

Another important difference between physical pathology and behavior pathology usually overlooked is that, in the case of most behavior disorders, it is not the underlying motivations that need to be altered or removed, but rather the ways in which the patient has learned to gratify his needs (Rotter, 1954). Thus, for example, if a patient displays deviant sexual behavior, the goal is not the removal of the underlying causes, i.e., sexual motivation, but rather the substitution of more socially approved instrumental and goal responses.

It might also be mentioned in passing, that, in the currently popular forms of psychotherapy, the role assumed by the therapist may bring him a good many direct or fantasied personal gratifications. In the course of treatment the patient may express considerable affection and admiration for the therapist, he may assign the therapist an omniscient status, and the reconstruction of the patient's history may be an intellectually stimulating activity. On the other hand, the methods derived from learning theory place the therapist in a less glamorous role, and this in itself may create some reluctance on the part of psychotherapists to part with the procedures currently in use.

Which of the two conceptual theories of personality—the psychodynamic or the social learning theory—is the more useful in generating effective procedures for the modification of human behavior remains to be demonstrated. While it is possible to



present logical arguments and impressive clinical evidence for the efficiency of either approach, the best proving ground is the laboratory.

In evaluating psychotherapeutic methods, the common practice is to compare changes in a treated group with those of a nontreated control group. One drawback of this approach is that, while it answers the question as to whether or not a particular treatment is more effective than no intervention in producing changes along specific dimensions for certain classes of patients, it does not

provide evidence concerning the relative effectiveness of alternative forms of psychotherapy.

It would be far more informative if, in future psychotherapy research, radically different forms of treatment were compared (King & Armistage, 1958; Rogers, 1959), since this approach would lead to a more rapid discarding of those of our cherished psychotherapeutic rituals that prove to be ineffective in, or even a handicap to, the successful treatment of emotional disorders.

## REFERENCES

- ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychol. Bull.*, 1957, **54**, 393-405.
- ALEXANDER, F. *Psychoanalysis and psychotherapy*. New York: Norton, 1956.
- ALEXANDER, F., & FRENCH, M. T. *Psychoanalytic therapy*. New York: Ronald, 1946.
- ATKINSON, RITA L. Paired-associate learning by schizophrenic and normal subjects under conditions of verbal reward and verbal punishment. Unpublished doctoral dissertation, Indiana University, 1957.
- AZKAN, N. H., & LINDSLEY, O. R. The reinforcement of cooperation between children. *J. abnorm. soc. Psychol.*, 1956, **52**, 100-102.
- BANDURA, A., & HUSTON, ALETHA, C. Identification as a process of incidental learning. *J. abnorm. soc. Psychol.*, in press.
- BANDURA, A., LIPSHER, D. H., & MILLER, PAULA, E. Psychotherapists' approach-avoidance reactions to patients' expressions of hostility. *J. consult. Psychol.*, 1960, **24**, 1-8.
- BANDURA, A., ROSS, DOROTHEA, & ROSS, SHEILA, A. Transmission of aggression through imitation of aggressive models. *J. abnorm. soc. Psychol.*, in press.
- BANDURA, A., & WALTERS, R. H. *Adolescent aggression*. New York: Ronald, 1959.
- CAIRNS, R. B. The influence of dependency-anxiety on the effectiveness of social reinforcers. Unpublished doctoral dissertation, Stanford University, 1959.
- CALVIN, A. D., CLIFFORD, L. T., CLIFFORD, B., BOLDEN, L., & HARVEY, J. Experimental validation of conditioned inhibition. *Psychol. Rep.*, 1956, **2**, 51-56.
- DAVIDSON, J. R., & DOUGLASS, E. Nocturnal enuresis: A special approach to treatment. *British med. J.*, 1950, **1**, 1345-1347.
- DITTES, J. E. Extinction during psychotherapy of GSR accompanying "embarrassing" statements. *J. abnorm. soc. Psychol.*, 1957, **54**, 187-191. (a)
- DITTES, J. E. Galvanic skin responses as a measure of patient's reaction to therapist's permissiveness. *J. abnorm. soc. Psychol.*, 1957, **55**, 295-303. (b)
- DOLLARD, J., AULD, F., & WHITE, A. M. *Steps in psychotherapy*. New York: Macmillan, 1954.
- DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
- DUNLAP, K. *Habits, their making and unmaking*. New York: Liveright, 1932.
- EDMONSON, B. W., & AMSEL, A. The effects of massing and distribution of extinction trials on the persistence of a fear-motivated instrumental response. *J. comp. physiol. Psychol.*, 1954, **47**, 117-123.
- ERIKSON, C. W. Unconscious processes. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. Nebraska Press, 1958.
- ESTES, W. K. An experimental study of punishment. *Psychol. Monogr.*, 1944, **57** (3, Whole No. 363).
- EYSENCK, H. J. *The dynamics of anxiety and hysteria*. New York: Praeger, 1957.
- FENICHEL, O. *Problems of psychoanalytic technique*. (Trans. by D. Brunswick) New York: Psychoanalytic Quarterly, 1941.
- FERSTER, C. B. Development of normal behavioral processes in autistic children. *Res. relat. Child.*, 1959, No. 9, 30. (Abstract)
- FISHMAN, H. C. A study of the efficiency of negative practice as a corrective for stammering. *J. speech Dis.*, 1937, **2**, 67-72.
- FRANK, J. D. The dynamics of the psychotherapeutic relationship. *Psychiatry*, 1959, **22**, 17-39.

- FROMM-REICHMANN, FRIEDA. *Principle of intensive psychotherapy*. Chicago: Univer. Chicago Press, 1950.
- HEINE, R. W. An investigation of the relationship between change in personality from psychotherapy as reported by patients and the factors seen by patients as producing change. Unpublished doctoral dissertation, University of Chicago, 1950.
- JONES, E. L. Exploration of experimental extinction and spontaneous recovery in stuttering. In W. Johnson (Ed.), *Stuttering in children and adults*. Minneapolis: Univer. Minnesota Press, 1955.
- JONES, H. G. The application of conditioning and learning techniques to the treatment of a psychiatric patient. *J. abnorm. soc. Psychol.*, 1956, **52**, 414-419.
- JONES, MARY C. The elimination of children's fears. *J. exp. Psychol.*, 1924, **7**, 382-390. (a)
- JONES, MARY C. A laboratory study of fear: The case of Peter. *J. genet. Psychol.*, 1924, **31**, 308-315. (b)
- KING, G. F., & ARMITAGE, S. G. An operant-interpersonal therapeutic approach to schizophrenics of extreme pathology. *Amer. Psychologist*, 1958, **13**, 358. (Abstract)
- KLEIN, MELANIE. *The psycho-analysis of children*. London: Hogarth, 1949.
- KRASNER, L. The use of generalized reinforcers in psychotherapy research. *Psychol. Rep.*, 1955, **1**, 19-25.
- KRASNER, L. Studies of the conditioning of verbal behavior. *Psychol. Bull.*, 1958, **55**, 148-170.
- LACEY, J. I., & SMITH, R. I. Conditioning and generalization of unconscious anxiety. *Science*, 1954, **120**, 1-8.
- LAZARUS, A. A., & RACHMAN, S. The use of systematic desensitization in psychotherapy. *S. Afr. med. J.*, 1957, **32**, 934-937.
- LEHNER, G. F. J. Negative practice as a psychotherapeutic technique. *J. gen. Psychol.*, 1954, **51**, 69-82.
- LINDSLEY, O. R. Operant conditioning methods applied to research in chronic schizophrenia. *Psychiat. res. Rep.*, 1956, **5**, 118-138.
- LIVERSEDGE, L. A., & SYLVESTER, J. D. Conditioning techniques in the treatment of writer's cramp. *Lancet*, 1955, **1**, 1147-1149.
- MARTIN, B., & KUBLY, DELORES. Results of treatment of enuresis by a conditioned response method. *J. consult. Psychol.*, 1955, **19**, 71-73.
- MAUSNER, B. Studies in social interaction: III. The effect of variation in one partner's prestige on the interaction of observer pairs. *J. appl. Psychol.*, 1953, **37**, 391-393.
- MAUSNER, B. The effect of one partner's success in a relevant task on the interaction of observer pairs. *J. abnorm. soc. Psychol.*, 1954, **49**, 557-560.
- MAX, L. W. Breaking up a homosexual fixation by the conditioned reaction technique: A case study. *Psychol. Bull.*, 1935, **32**, 734.
- MEISSNER, J. H. The relationship between voluntary nonfluency and stuttering. *J. speech Dis.*, 1946, **11**, 13-33.
- MEYER, V. The treatment of two phobic patients on the basis of learning principles: Case report. *J. abnorm. soc. Psychol.*, 1957, **55**, 261-266.
- MILLER, N. E., & DOLLARD, J. *Social learning and imitation*. New Haven: Yale Univer. Press, 1941.
- MORGAN, J. J. B., & WITMER, F. J. The treatment of enuresis by the conditioned reaction technique. *J. genet. Psychol.*, 1939, **55**, 59-65.
- MOWRER, O. H. *Learning theory and personality dynamics*. New York: Ronald, 1950.
- MOWRER, O. H., & MOWRER, W. M. Enuresis—a method for its study and treatment. *Amer. J. Orthopsychiat.*, 1938, **8**, 436-459.
- MURRAY, E. J. The content-analysis method of studying psychotherapy. *Psychol. Monogr.*, 1956, **70** (13, Whole No. 420).
- MUSSEN, P., & DISTLER, L. M. Masculinity, identification, and father-son relationships. *J. abnorm. soc. Psychol.*, 1959, **59**, 350-356.
- MUSSEN, P., & DISTLER, L. M. Child-rearing antecedents of masculine identification in kindergarten boys. *Child Developm.*, 1960, **31**, 89-100.
- PAYNE, D. E., & MUSSEN, P. H. Parent-child relationships and father identification among adolescent boys. *J. abnorm. soc. Psychol.*, 1956, **52**, 358-362.
- PETERS, H. N. Multiple choice learning in the chronic schizophrenic. *J. clin. Psychol.*, 1953, **9**, 328-333.
- PETERS, H. N., & JENKINS, R. L. Improvement of chronic schizophrenic patients with guided problem-solving motivated by hunger. *Psychiat. Quart. Suppl.*, 1954, **28**, 84-101.
- RACHMAN, S. The treatment of anxiety and phobic reactions by systematic desensitization psychotherapy. *J. abnorm. soc. Psychol.*, 1959, **58**, 259-263.
- RAYMOND, M. S. Case of fetishism treated by aversion therapy. *Brit. med. J.*, 1956, **2**, 854-857.
- RAZAN, G. Stimulus generalization of conditioned responses. *Psychol. Bull.*, 1949, **46**, 337-365.
- ROBINSON, NANCY M. Paired-associate learning by schizophrenic subjects under conditions of personal and impersonal reward

- and punishment. Unpublished doctoral dissertation, Stanford University, 1957.
- ROGERS, C. R. *Client-centered therapy*. Boston: Houghton Mifflin, 1951.
- ROGERS, C. R. Group discussion: Problems of controls. In E. H. Rubinstein & M. B. Parloff (Eds.), *Research in psychotherapy*. Washington, D. C.: American Psychological Association, 1959.
- ROGERS, J. M. Operant conditioning in a quasi-therapy setting. *J. abnorm. soc. Psychol.*, 1960, 60, 247-252.
- ROSENTHAL, D. Changes in some moral values following psychotherapy. *J. consult. Psychol.*, 1955, 19, 431-436.
- ROTTER, J. B. *Social learning and clinical psychology*. Englewood Cliffs, N. J.: Prentice-Hall, 1954.
- RUBENSTEIN, E. A., & LORR, M. A comparison of terminators and remainers in outpatient psychotherapy. *J. clin. Psychol.*, 1956, 12, 345-349.
- RUTHERFORD, B. R. The use of negative practice in speech therapy with children handicapped by cerebral palsy, athetoid type. *J. speech Dis.*, 1940, 5, 259-264.
- SALZINGER, K. Experimental manipulation of verbal behavior: A review. *J. gen. Psychol.*, 1959, 61, 65-94.
- SARASON, BARBARA R. The effects of verbally conditioned response classes on post-conditioning tasks. *Dissertation Abstr.*, 1957, 12, 679.
- SCHMIDBERG, MELITTA. Psychotherapy of juvenile delinquents. *Int. ment. hlth. res. Newsltr.*, 1959, 1, 1-2.
- SEARS, PAULINE S. Child-rearing factors related to playing of sex-typed roles. *Amer. Psychologist*, 1953, 8, 431. (Abstract)
- SEARS, R. R. Identification as a form of behavioral development. In D. B. Harris (Ed.), *The concept of development: An issue in the study of human behavior*. Minneapolis: Univ. Minnesota Press, 1957.
- SHEEHAN, J. G. The modification of stuttering through non-reinforcement. *J. abnorm. soc. Psychol.*, 1951, 46, 51-63.
- SHEEHAN, J. G., & VOAS, R. B. Stuttering as conflict: I. Comparison of therapy techniques involving approach and avoidance. *J. speech Dis.*, 1957, 22, 714-723.
- SHOBEN, E. J. Psychotherapy as a problem in learning theory. *Psychol. Bull.*, 1949, 46, 366-392.
- SKINNER, B. F. *Science and human behavior*. New York: Macmillan, 1953.
- SKINNER, B. F. Some issues concerning the control of human behavior. *Science*, 1956, 124, 1057-1066. (a)
- SKINNER, B. F. What is psychotic behavior? In, *Theory and treatment of psychosis: Some newer aspects*. St. Louis: Washington Univ. Stud., 1956. (b)
- SOLOMON, R. L., KAMIN, L. J., & WYNNE, L. C. Traumatic avoidance learning: The outcomes of several extinction procedures with dogs. *J. abnorm. soc. Psychol.*, 1953, 48, 291-302.
- SULLIVAN, H. S. *The interpersonal theory of psychiatry*. New York: Norton, 1953.
- THIRMAN, J. Conditioned-reflex treatment of alcoholism. *New Engl. J. Med.*, 1949, 241, 368-370, 406-410.
- THOMPSON, G. N., & BIELINSKI, B. Improvement in psychosis following conditioned reflex treatment in alcoholism. *J. nerv. ment. Dis.*, 1953, 117, 537-543.
- TILTON, J. R. The use of instrumental motor and verbal learning techniques in the treatment of chronic schizophrenics. Unpublished doctoral dissertation, Michigan State University, 1956.
- ULLMAN, L. P., KRASNER, L., & COLLINS, Beverly J. Modification of behavior in group therapy associated with verbal conditioning. *J. abnorm. soc. Psychol.*, in press.
- VOEGTLEN, W. L. The treatment of alcoholism by establishing a conditioned reflex. *Amer. J. med. Sci.*, 1940, 119, 802-810.
- WALLACE, J. A. The treatment of alcoholics by the conditioned reflex method. *J. Tenn. Med. Ass.*, 1949, 42, 125-128.
- WEIDE, T. N. Conditioning and generalization of the use of affect-relevant words. Unpublished doctoral dissertation, Stanford University, 1959.
- WHITING, J. W. M. The research program of the Laboratory of Human Development: The development of self-control. Cambridge: Harvard University, 1954. (Mimeo)
- WILLIAMS, C. D. The elimination of tantrum behaviors by extinction procedures. *J. abnorm. soc. Psychol.*, 1959, 59, 269.
- WILLIAMS, R. I. Verbal conditioning in psychotherapy. *Amer. Psychologist*, 1959, 14, 388. (Abstract)
- WOLFE, J. Reciprocal inhibition as the main basis of psychotherapeutic effects. *AMA Arch. Neurol. Psychiat.*, 1954, 72, 205-226.
- WOLFE, J. *Psychotherapy by reciprocal inhibition*. Stanford: Stanford Univ. Press, 1958.
- WOLFE, J. Psychotherapy based on the principle of reciprocal inhibition. In A. Burton (Ed.), *Case studies in counseling and psychotherapy*. Englewood Cliffs, N. J.: Prentice-Hall, 1959.
- YATES, A. J. The application of learning theory to the treatment of tics. *J. abnorm. soc. Psychol.*, 1958, 56, 175-182.

(Received January 21, 1960)

## THEORY OF SITUATIONAL, INSTRUMENT, SECOND ORDER, AND REFRACTION FACTORS IN PERSONALITY STRUCTURE RESEARCH

RAYMOND B. CATTELL

*University of Illinois*

Exploration of personality by multivariate experimental methods, as a means of objectively determining personality structure, has revealed, on the one hand, an array of stable, meaningful, cross-checking structures (Cattell, 1946, 1957; French, 1953), and on the other, some baffling inconsistencies. The latter have recently been pointed out by Becker (1960), apparently in criticism of the present writer's personality theory, but have been known for several years, and were, in fact, first brought to light by Cattell and Saunders (1950). Nevertheless, Becker does a service to advertise these facts; for psychologists have greatly neglected the solution of the problems revealed in this field.

The present writer's theoretical position is that it is conceptually correct to speak of the same unique source trait, e.g., cyclothymia-schizothymia, anxiety, ego-strength, surgency-desurgency, as something expressing itself (in terms of recognizable, replicable factor patterns) across all three possible media of experimental observation. That is to say, the same influence should appear in L data (life record, behavior *in situ*), Q data (questionnaire, consulting room, verbal self-evaluation), and T data (objective, laboratory, miniature situational, non-self-evaluative test performances).

In the article (Becker, 1960) to which I reply the fact that the actual correlation between the L-data and

Q-data estimates of what are apparently equivalents in the two media, sometimes falls far short of perfection, is accepted as disproof of this theory. This theoretical conclusion is unsubtle; and the thesis of my reply is that countless threads of evidence contribute to the view that the same abstract personality source trait commonly operates across different media. However, certain "perturbations" have to be recognized which prevent the simple relation appearing on the surface, and these need to be taken into account in understanding psychological measurement generally.

In this area of scientific investigation, Becker has not asked the right question. Unexpected, but systematically evaluated perturbations of existing laws have often led to new discoveries, not so much by rejecting a law as by extending it, e.g., in astronomy in the discovery of Neptune through observed perturbations in the expected orbit of Uranus. So here, it is argued that there is no reason to abandon the notion of unitary source traits (Cattell, 1946) but that one must recognize certain new concepts, which we have introduced under the terms situational, instrument, and refraction factors. These are supported partly by marshaling existing evidence, but also by experiments undertaken ad hoc, but which, through an editorial veto on space to reply, have been reported in a separate publication (Cattell, 1960).

#### THE DEFINITION OF INSTRUMENT FACTORS

The first and major source of perturbation in transmedia factor matching arises from what may be called *instrument factors*. Apparently, the first explicit recognition and demonstration of an instrument factor occurred in a structural analysis of a very widely selected set of objective personality tests, by Cattell and Gruen (1955), where a factor appeared literally produced by diurnal variations of sensitivity of a brass instrument (GSR). This purely instrumental influence created a factor by throwing common variance into all types of personality measures in which it was used. Such factors have appeared since in publications by Holzmann and Bitterman (1956), F. L. Damarin, D. T. Campbell, and L. Berwyn (unpublished), and several other unpublished studies known to the writer. Indeed, wherever questionnaire variables are mixed with ratings, attitude scales with questionnaires, or, sometimes, even one type of answer form with another, one or more factors may generally be found covering *all variables having formal similarity*.

The difficulty factors of Wherry and Gaylord (1944), and Dingman (1958), should definitely be regarded as a subspecies of instrument factor. Recently, in a study of the Music Preference Test of Personality (Cattell & Anderson, 1953) by Mayeske (1961) an instrument factor appeared even separating all items resting on one form of musical recording from those based on another technique. Instrument factors have become better understood in the last couple of years through extensive studies of their appearance in objective motivation structure analyses (Cattell,

Radcliffe, & Sweney, 1960; R. B. Cattell & J. Horn, unpublished). There they appear as "vehicle factors" covering all objective devices using the same vehicle, e.g., information, autism, for the objective measurement of motivation strength. In this, and many similar contexts, it has been shown that instrument factors can be fairly clearly eliminated by ipsative scoring (R. B. Cattell & J. Horn, unpublished, see Table 1).

Before proceeding beyond this introduction by illustrations, to a more comprehensive definition of the concept of instrument factor, it is desirable, however, to make clear which peripheral factors are *not* to be included. This can be done most compactly by Figure 1, presenting a hierarchy which will be clear to multivariate experimentalists. Incidentally, the term "artifacts" is due to Roberts (1959), and has been sharpened by additional conditions here to make their separation from instrument factors cleaner.

The justification for the labels of the three forms of "perturbing" factors reproducible across experiments (matrices) will be given as we proceed. Concentrating first on instrument factors, let us note that they are definable, initially, only in terms of intention and perspective. Later, the definition can be made more satisfactory as we develop precise concepts indicating various universes of variables. For a quality which persists across the differences of content of a series of opinionnaires of similar form, and which perhaps consists of response to a particular form inherent in this instrument, though irrelevant to the content interest of the experimenter may yet represent behavior dependent on a real per-



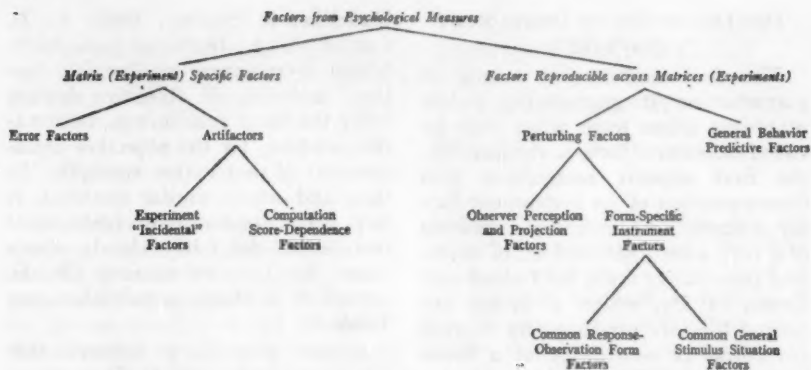


FIG. 1. The place of instrument factors in a taxonomy of factors.

sonality trait. For example, what comes as an instrument factor covering the variables of similar form,  $a_1 \dots a_n$ , may well load (when  $a_1 \dots a_n$  are condensed to a single variable  $a$ , set in the new context of variables  $b, c, d$ , etc.) some important general personality factor.<sup>1</sup>

There is thus a sense in which an instrument factor is a matter of perspective, i.e., of one's starting point

<sup>1</sup> Incidentally, it is the failure to recognize this perspective which, in the present writer's opinion, has made so much recent work on response sets a rather uneconomical use of psychological research time. Whereas educational psychometrists during the late 1950s "discovered," in their opinionnaire tests, response sets (Cronbach, 1950), social desirability sets (Edwards, 1957), extremity of response sets (Berg, 1955), and acquiescence—tendency to agree, yes-vs.-no (Messick & Jackson, 1961)—these had already been employed by designers of objective personality tests in the late 1940s and early 1950s (Cattell, 1946; Cattell & Gruen, 1955). In the context of broader personality theories, and varied behavioral measures involved, it had already become clear that what itemetrists, without knowledge of the literature in this area, later treated merely as "flaws" in their paper-and-pencil tests, were actually expressions of well defined personality factors, e. g., anxiety or UI 24, comention or UI 20, super-ego rigidity or UI 29, as well as UI 31 (Cattell, 1957).

and of the plane of experience from which one chooses the majority of one's tests. In this sense, just as dirt is only "matter in the wrong place," so an instrument factor is only "variance where we didn't expect it or don't want it." When we are measuring personality by questionnaire we obviously do not want *each and all* of the diverse personality dimensions included to be contaminated by what might be called a "generalized specific," i.e., a specific to questionnaires. And the fact that that specific may, indeed, be something more than a trivial specific, but an expression of a single important personality factor spread over and contaminating all the alleged diverse personality measures, does not make the measurement harsh any more acceptable!

When more progress has been made toward a systematic taxonomy of tests, on some such objective basis as that worked out by Cattell and Warburton (in press), it would become possible to set up also a relatively objective classification of instrument factors, according to the types of personality approach to which they are tangential. For "form" and "content" are quite

subjective categories, and, in any case, by no means exhaust the possible planes of experiment to which instrument factors can be orthogonally intrusive. For the time being, however, we must take a relativistic position, and one centered in "content." On this basis we shall contingently define an instrument factor as any uniquely (simple structure) rotated factor which covers a whole set of diverse variables having formal resemblance in presentation, mode of permitted response, or scoring, and which does not extend to tests of the same psychological content when couched in other modes of formal presentation, response, etc.

#### THEORY OF SOURCES OF PERTURBATION AFFECTING TRAIT ALIGNMENT

It should be noted that there are two distinct, though related senses in which a source trait can be said to be the same or not the same in two different media:

1. An estimate of the factor from the variables in one medium may correlate less than unity with its estimate from variables in another medium, even when attenuation-corrected for (a) unreliability of measurement, and (b) imperfection of estimate.

2. It may not be possible to discover a trait, when factoring both media together, which has simple structure across both media and also possession of the hypothesized, similar-meaning salient loadings in both media. (Whether one also means that the simple structure position in one medium will not project into the other we shall discuss below.)

Becker has been concerned with the first of these, denying alignment without first checking that Corrections *a* and *b* could not restore the

correlation to unity. In any case, the second meaning is more important. If unity in this sense holds, personality theory is profoundly simplified, and it is only a matter of the mechanics of statistics to produce weighted measures from the two media that *will* approach a correlation of unity.<sup>2</sup>

In the larger collation of data and new experimental work (Cattell, 1960), from which the present article abstracts, it has been shown that the presence of unrecognized instrument factors in the two media will prevent alignment either in Sense 1 or 2, unless special new techniques are used. Before devoting a section to closer inspection of this result, however, it is desirable to set out a clear theory about more general sources of perturbation. For, in principle, one can see that there are some six possible origins of the failure to find a one-to-one alignment of primary personality factors measured in one medium with those measured in another. Some of these will produce instrument factors; others will contribute to other kinds of nonalignment to be described.

#### Sources of Nonalignment

*Human transmission (perception, evaluation, projection, memory) of score values.* Largely this means rating and self-rating (L and Q data). This is too subtle and complex a field—hitherto handled too simply in terms

<sup>2</sup> Such a procedure should be sharply distinguished from what Becker (1960) appears to advocate, and describes Gough as doing, namely, to force a Q scale to align itself with an L factor by assiduous item selection. Any such procedure contributes nothing to our knowledge of structure, but only hides the problem. If it succeeds, and if our theory is correct that L-data factors are the most heavily contaminated of any medium with irrelevant factors, this is forcing a poorly oriented measure to agree with a still poorer one.

of "halo"—for the present abstract summary to be illustrated in available space (see Cattell, 1960). Theoretically, the pattern of correlations, and therefore of obtained factors, could be distorted by, and only by, properties of the individual and his relation to the recorder which affect the recording of *all* his behavior variables, and by properties of the perceiving recorder. The former can be divided into (a) value relationships, of which liking-disliking (a constituent in halo) is only one; and (b) perspicacity or visibility effects, e.g., extraversion making the ratee more known, position effects making certain behaviors more clear. The latter can be divided into projections of (a) stereotypes or cultural clichés,<sup>3</sup> and (b) refraction factors, discussed below, peculiar to one medium. In all the "perceiving recorder effects" a correlation is produced by "projection" of a (perhaps quite unconscious) conviction that certain variables go together. Some of these may produce typical instrument factors, uniformly and about equally loading all variables in the medium; but others may load only *some* variables, producing what are perhaps best described as "perception-evaluation" projection factors, and which are not true instrument factors.

*Communality of variables in respect to some trait required for handling a similar formal performance in all of them (or for registering in an observation situation).* This is essentially one of the two main sources (see following paragraph) of instrument factors only. The countless possibili-

ties may be illustrated by e.g. the use of 30 scales in all of which the score (in one direction or the other) depends on an ability to read, or on information or skill of expression, or tendency to say yes rather than no, etc.

*Communality of variables in respect to scoring or scaling applied after administration.* Quite apart from common demands on the subject's actual performance as in the previous paragraph, anything in the formal *scoring* procedure which tends to give similar sigmas, and skewedness (and in some cases means) throughout one class of tests will tend to create higher correlation among them and a common factor. That is to say, if the matrix of correlations of tests  $a_1$  through  $a_n$  were just the same, on a rank formula, as that for  $b_1$  through  $b_n$ , but if all the  $a$ 's, on the one hand, and all  $b$ 's, on the other, have similar distribution, then basing the matrix afresh on a product-moment formula will tend to give an instrument factor for the  $a$ 's and/or the  $b$ 's separately.

*Coincidence of different global stimulus situations with different test media administrations.* If a person answered one set of questionnaires in private, and another orally and publicly (which is akin to the interview or behavior rating situation), we should expect real differences in response due to the actual stimulus situation, covering the occasion on which all items of one test were answered, being different from that covering the other test-taking setting. A priori this could create both an instrument factor, conterminous with each medium-situation, and also a change in loading of the same items on the same personality factors in the two situations.

*Habitual broad area differences in actual trait development and expression.* Among children, for example,

<sup>3</sup> Since sociologists have ruined "stereotype," by applying it equally to a widespread concept which either (a) does or (b) does not, correspond to statistical reality, I suggest "cultural cliché" explicitly for a widespread cultural concept which is significantly different from any externally existing pattern.

we should expect the particular behavior variables representing, say, the dominance factor, to be expressed to different degrees in the home environment and in the school environment. This is analogous to the point in the above paragraph, except that the influence is expressly conceived not to lie in the temporary measurement situation itself, but in the prolonged life situation, leading to real differences of actual habit strength, i.e., of the trait itself. Factor analytically, this might produce a home dominance factor and a school dominance factor, representing the relative impact of home and school, respectively, or alternatively, one factor modified by two other factors, each peculiar to one broad area. If the former proves to be more characteristic, then we can confidently predict that the two first-order factors will correlate highly and yield a single second-order dominance factor. Even if the former is true it would be possible, in a rough factoring to perceive the structure as that of a home and school instrument factor (as in the second possibility) but psychologically, the interpretation, if the proper structure is obtained, would now be different from an instrument factor effect. The area differences would then be interpreted as *real* structure differences, and the concept of a single dominance trait would be discovered and justified *only* at the second-order factor level.

*Differences among media in density of representation of variables.* If in sampling variables in the ability field an experimenter accidentally took one variable for each of Thurstone's primary abilities and factored, he would obtain, straightaway, i.e., as a first-order factor, that general ability factor which, in any "dense" representation of variables, appears only as a second-order factor (Thurstone,

1938). This concept of density of variable representation has been developed further elsewhere (Cattell, 1957, pp. 808-817), but it is easy to see that if there were really large differences of density unrecognized between media we should obtain no correlational alignment of the primaries in the two fields. Only on exploring the second order would the possibility arise of discovering that a second order in one medium is the same as a first order in the other.

Actually, as soon as systematic exploration of second-order structure in questionnaires reached to six factors (Cattell, 1957; Cattell & Scheier, 1961; Cattell & Warburton, in press), it became evident that four second-order *questionnaire* factors aligned with four first-order *objective test* factors (UI 19, 20, 24, and 32); and in two of these, UI 24 (anxiety) and UI 32 (extraversion), the agreement is perfect within small limits of experimental error. An instance from a different realm, but amounting to a correlation of only 0.80 between the two media, exists in Tollefson's demonstration (1961) that the second-order extraversion factor in the questionnaire is a first-order factor in the Humor Test of Personality. These alignments (from the earlier, 1954-1957, publications above) are not mentioned in Becker's article (1960), perhaps because his comments are all on L- and Q- (rather than T-) data alignments. But the findings are highly relevant as showing that there does exist a corner of the intermedia jigsaw puzzle which is beginning to fit in place. These five experimental instances alone are surely sufficient to encourage us in that rejection of nihilism which this article undertakes.

To risk a prediction in the little explored field of "density," one might judge that variables in Q data

will prove somewhat more "dense" than L data. But substantially, as the above evidence shows, one can conclude only that variables as commonly chosen are much more dense in Q than T data. This is understandable; e.g., in the T-data anxiety factor, we test startle response by a single cold pressor test (Cattell & Scheier, 1961) whereas in most anxiety questionnaires there are a dozen items asking in different ways how easily the person startles. Cronbach (1960), Comrey, and others who criticize low homogeneity when reviewing factor scales, are perhaps unwittingly driving their flocks toward the more serious danger of using personality scales heavily loaded in spurious "specific" variance of this latter kind, instead of watching that their scales deal with personality factors having broad psychological relevance and effectiveness.

If the above search for sources of perturbation has been truly exhaustive, our summary must include three other forms of distortion besides instrument factors, constituting four in all, as follows (beginning with instrument factors):

1. Test instrument factors, including common test form (response-observation-score) factors, and common test general stimulus situation factors.

2. Modification of actual trait by influences peculiar to one area of expression, producing primaries for each area and requiring conceptual unity to be sought at a higher order level.

3. Difference of density of representation of variables, as commonly unconsciously chosen by experimenters, in their different media, resulting in a higher order in one medium matching a lower order in another.

4. Perception-evaluation or projection factors, which trespass on the

variance of the variables used to estimate personality factors, *not* by uniformly loading all in one medium (as does an instrument factor) but having each a characteristic form, and, when restricted to one medium, having the properties of refraction factors described below.

#### THE PRACTICAL PROBLEM OF REACHING PERSONALITY STRUCTURE DESPITE DISTORTIONS

If the above theoretical analysis is correct the manifest correlational picture of personality structures will be less like Whistler's portrait of his mother than the cubist's rendering of the same, fractured into surprising new supernumerary planes and facets. To translate from the latter to the former, it is necessary that research, first, check the hypotheses about the forms of distortion at work and, second, find experimental and statistical means for isolating and setting aside these various perturbing influences.

One cannot do more than glance at these tasks here. As to the first, our initial examination of data shows definitely that form-specific instrument factors exist, while my colleagues and I have also begun to give evidence for the Sources 2, 3, and 4. The source of nonalignment labeled 2—local area modification of real traits—has been more fully illustrated elsewhere (Cattell, 1960) but must be left to others systematically to investigate. Source 3, changing density with changing medium, has already been substantiated.

As to the second task—segregating the distorting influences to arrive at essential structure—the unraveling of Effects 2 and 3 above is straightforward, by second-order factoring, though the possibility has been mooted above that Source 2 could



produce two instrument factors, beyond a single first-order factor, instead of two first orders resolving into a second.

Setting 2 and 3 aside, therefore, we shall devote the present section to unraveling the effect of instrument factors, 1 above, and the following section to perception-evaluation-project phenomena, 4 above.

The special experiments with instrument factors described elsewhere (Cattell, 1960) proceeded first to find what happens when one factors correlation matrices derived from known, numerically stated factor models, and secondly, to experiment with varieties of solution in actual psychological data where the existence and boundaries of an instrument factor were well known beforehand. These experiments showed that:

1. Where the instrument factor covers *all* variables, i.e., where they are not embedded in a larger matrix, with other media to constitute a hyperplane and determine unique rotation, the typical investigator and procedure will not find or be aware of the instrument factor.

2. If the instrument factor is not found then either: (a) the correlations among the primaries will be distorted (if it is positive on all and they are all positively correlated, it will increase their correlations); or, (b) the simple structure which really exists among the primaries will not be found, or found only in very impaired form. Commonly *b* will predominate, but both will operate.

After this demonstration of the effect of an instrument factor in a single medium we proceeded to models and real instances containing blocks of variables uniformly from each of two or three media. Herein each medium was covered by *one* instrument factor but where *true* personality factors existed in the sense of

having a simple structure position with salient loadings on variables of similar meanings in *both* media. Here it was shown:

1. If one obtains the best possible simple structure (perhaps imperfect because of mixed-in instrument factor) among variables separately in each medium, the same simple structures cannot usually be found when the media are put together.

2. One reason for this is that if one projects the simple structure position satisfactorily obtained in one medium into the second,<sup>4</sup> it definitely does not give simple structure within the second.

3. If, however, one first admits the existence of, and locates by simple structure in the combined matrix, the instrument factors (which can now have determinate hyperplanes), then the true personality factors, operating across both media, can be located (in blind simple structure rotation). A successful example of this in real data—objective motivation measurement (R. B. Cattell & J. Horn, unpublished)—is shown in Table 1 here, and in other models elsewhere (Cattell, 1960). Our ignorance of this principle in 1948 was presumably responsible for the chaotic outcome of the first extensive transmedium factor analyses (Cattell & Saunders, 1950, 1955).

Incidentally, it will be obvious that missing the instrument factor, failing to rotate it correctly if one does not miss it, and encountering the subsequent distortion are due respectively to (a) the lack of a test for factor extraction that will decide,

<sup>4</sup> This cannot be done, of course, simply by applying the same discovered transformation ( $\lambda$ ) matrix to the centroids, because the latter begin at different positions. One first discovers by the Procrustes program the  $\lambda$  most nearly reproducing the first medium simple structure from the joint medium centroid.

TABLE 1  
PSYCHOLOGICAL AND INSTRUMENT FACTORS AS FOUND IN OBJECTIVE,  
DYNAMIC TRAIT SIMPLE STRUCTURE

Attitude Variable and Device Measurement	Factor Matrix				
	Psychological Factors			Instrument Factors	
	Escape Erg	Sentiment to Parents	Self-Sentiment	Information Device Factor	Autism Device Factor
1 Desire for good self-control. Information measure	00	-02	26	54	03
2 Wish to know oneself. Information measure	03	-05	31	27	19
3 Wish to never to become insane. Information measure	-06	12	22	43	04
4 Readiness to turn to parents for help. Information measure	-02	35	09	28	-01
5 Feeling proud of one's parents. Information measure	-06	28	-01	24	01
6 <sup>a</sup> Desire to avoid fatal disease and accidents. Information measure	16	04	13	65	-02
7 <sup>a</sup> Wish to get protection from A bomb. Information measure	14	-08	03	14	-05
8 Desire for good self-control. Autism measure	01	-04	30	02	22
9 Wish to know oneself. Autism measure	-08	07	37	-01	31
10 Wish never to become insane. Autism measure	00	-01	16	00	25
11 Readiness to turn to parents for help. Autism measure	-08	18	09	-08	42
12 Feeling proud of one's parents. Autism measure	-03	14	01	06	14
13 <sup>a</sup> Desire to avoid fatal disease and accidents. Autism measure	26	20	01	04	17
14 <sup>a</sup> Wish to get protection from A bomb. Autism measure	23	13	09	15	10

Note.—The theoretically required salients to define the factors are boxed in, and except for two values at the bottom of the parental sentiment factor column, the salients are high (above .09) where, and only where, they are theoretically required to be.

<sup>a</sup> Attitudes 13 and 14 are the same as 6 and 7, but in a different medium, and similarly, for the other cross-media personality factors.

to within less than an error of two or three factors, how many should be extracted; (b) having no variables from other media to give a hyperplane for it; and (c) the variance that should have been in the instrument factor being pushed into the personality factors, destroying the clarity of their hyperplanes. The remedy which worked in the above cases was to give good technical attention to these issues.

#### ON ISOLATING TRANSMEDIUM PERSONALITY FACTORS AND REFRACTION FACTORS

Our final step consisted in returning to the actual L and Q data from which Becker infers that personality factors are unmatchable across media, and showing that when examined by more penetrating concepts, as above, uniquely determinate, psychologically meaningful, factor patterns appear, expressing themselves appropriately in both media for each factor. This has theoretical interest in giving additional substance to Point 3 above, by introducing the no-

tion of refraction factors, and in producing some order in that L-Q frontier which has hitherto been the most hopelessly obscure of the transmedia relationships. Nevertheless, this approach does no more than reveal *some* order, and at the same time opens the door on a lot of problems, particularly in the field of behavior rating, which will now demand systematic investigations.

It is not easy to find in any published study of the past 20 years (ever since personality structure research began in earnest) an experiment really adequate in reaching the technical conditions necessary to get anywhere on this question. One needs, among other things, an experiment: (a) on a sufficient sample for sampling errors not to be intrusive; (b) where the subjects had a long testing period in which they were simultaneously rated *in situ* and subjected to questionnaires, comprehensive, reliable, and valid enough to define several factors clearly; (c) where ratings and questionnaire variables were strategically chosen to

represent psychologically *familiar* factors, already vouched for by earlier researches; and (d) where ratings were carried out by peers and under the requisite conditions described elsewhere (Cattell, 1946, 1957). Probably the most satisfactory data available is that in which the experimental work was broadly conceived and painstakingly carried out by Coan, on 7.8-year-old children (Cattell & Coan, 1957, 1958). It suffers only with respect to *d*, in that ratings were made by teachers instead of peers, and perhaps in reduced homogeneity of sample through equal inclusion of boys and girls.

Taking the data of this experiment we find that 24 rating variables have already been factored and blindly rotated into 12 very definite simple structure factors, each represented by two markers (see Table 5 in Cattell, 1960). Similarly, 24 variables in Q data, each consisting of a scale of about eight items, have been resolved as 12 well known simple structure factors, each marked essentially by two salient variables. However, on psychological inspection of these resolutions, the hypothetical position was taken that only 10 of the 12 factors were common to the two matrices, the remaining 4 being special, 2 to each matrix.

The two sets of 24 variables were now combined and intercorrelated in a cross-medium, L-Q matrix of 48 variables, which, by Tucker's test, yielded 16 factors. (With the hypothesis of matching, above, one would expect 14, but it is usual to find some new factor created by the mixture when two matrices are pooled.) The structure of this new factor space proved to be complex. Projection of simple structure obtained in one into the other, as described earlier (Footnote 4), would not yield a good combined simple

structure. Attempts to force simple structure by varimax, oblimax, or other "analytical" programs failed because these rigid programs could not recognize and uniquely rotate the instrument factors, which, on the basis of the above principles and findings, we knew must be present. Only a patient and comprehensive exploratory visual rotation (aided by the photographic Rotoplot program on Illiac), over 22 rotations, yielded a position of such stability that one could repeatedly return to it. In reaching this position we found that the hyperplanes in the data were noticeably a little broader (about  $\pm .13$  instead of  $\pm .10$ ) than those existing in one medium alone.

On examining the solution, set out in Table 2,<sup>5</sup> we found that we had essentially an instrument factor for L data and another for Q data (not set out at the end of the matrix, but marked  $1n_L$  and  $1n_Q$ , in Table 2). There are also two other factors, which we would guess might be projected "clichés," numbered 13 and 16. The interesting fact is that when this debris is set aside, patterns for the well known personality dimensions C (Ego strength), D (Excitability), F (Surgency), and H (Parmia), appear, with the appropriate four markers (2L and 2Q) on each, though the hyperplanes are pierced by one or two random appreciable loadings on other factors. (Counting within  $\pm .13$  they reach acceptable percent-

<sup>5</sup> The matrix containing the correlations among factors, the lambda matrix, and the centroid for Table 2 have been deposited with the American Documentation Institute. Order Document No. 6570 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress; Washington 25, D. C., remitting in advance \$1.75 for microfilm or \$2.50 for photocopies. Make checks payable to: Chief, Photoduplication Service, Library of Congress.

TABLE 2  
SIMPLE STRUCTURE ROTATION OF COMBINED L AND Q DATA, WITH REGARD FOR INSTRUMENT FACTORS  
(Primary Factor Pattern)

	L	C	D	F	H	InL	Inq	Aq	GL	Gq	Jq	CL	Oq	13	Qq	Dq	16
A	+20	+03	-03	+05	+06	+44	+09	-10	-07	-16	+04	-20	+06	-48	+02	-01	-25
C	<u>+45</u>	<u>+07</u>	<u>+13</u>	-15	-03	+04	-01	-19	-04	+05	+05	-64	-11	-16	-08	+00	-03
D	+13	<u>+45</u>	-08	+00	-12	+67	-03	-11	-10	-03	+15	+02	+04	+07	-03	-10	-10
E	+43	-31	+30	-08	+30	+43	+03	+09	-10	+01	-09	-16	+21	+04	+04	-04	+22
F	+39	-05	-09	<u>+50</u>	-09	-22	-01	+10	+06	+07	+04	+14	+08	+07	-01	-07	-25
G	-19	-09	-03	+01	+08	-19	+01	-03	<u>+66</u>	+00	-01	-01	-11	+05	-04	-07	+00
H	+10	-03	-11	-07	<u>+48</u>	-11	-06	-10	<u>+50</u>	+08	-01	-01	-04	-22	+06	+04	-07
I	-14	+10	-03	-01	-05	-56	-21	-10	-13	-11	+00	+06	+40	-07	+00	-10	+19
J	-20	+31	+19	-01	+09	+45	+09	-35	-04	+00	-08	+00	-33	-04	-09	-10	+05
L	+00	+19	+06	+00	+14	+60	-02	+00	-03	+09	+05	-12	-05	+02	-04	-09	+08
M	-08	-12	-11	+01	-14	-12	-03	+16	+06	-01	-01	+10	+10	-53	+08	+13	-09
O	-10	-09	-09	+07	-14	-11	-18	+16	-18	+12	-09	-02	-10	-53	+08	+13	+15
A	+03	-28	-12	+15	+53	+18	+15	+06	+00	+11	+04	-02	-08	-10	+03	-19	-02
C	<u>+49</u>	<u>+03</u>	-20	+19	+19	-22	+03	+09	+07	-01	+08	-56	+08	+11	+10	+04	-06
D	+00	<u>+20</u>	<u>+05</u>	-00	-14	+84	+00	-05	-10	+09	+04	+12	-06	+07	-07	-13	-13
E	+00	+03	+03	+01	+48	+71	+02	+07	+19	+03	+05	+10	+18	-10	+08	-19	-01
F	-13	+12	+12	<u>+46</u>	+27	-21	-10	+06	-65	-04	-09	+10	-01	-02	+02	-07	+08
G	+07	+07	+07	+07	+25	-19	+07	-01	<u>+76</u>	-04	-05	-08	-04	+13	+00	+06	+11
H	-07	-09	-09	+04	<u>+55</u>	-21	+02	+06	+22	-07	-02	-04	-08	+08	-08	-23	+01
I	+32	-01	-01	-07	+09	+05	+12	+05	-22	-03	-05	-07	+84	+13	+01	-03	+05
J	-06	+11	+06	-04	-41	+84	+07	+04	+07	+01	-05	-10	-08	-02	+02	-27	+10
L	+10	+06	+06	-11	+00	+67	+03	+07	+10	+00	-13	-35	-06	-03	-02	-24	-14
M	+12	+01	+01	-07	+44	+40	+03	-05	+37	+04	+04	+12	+04	-42	-01	+27	+01
O	-03	-22	-22	+07	-48	-54	+07	-14	+13	+07	-07	+20	-00	-16	-01	+10	+30
A	-03	+06	+06	-19	+06	+10	+01	<u>+58</u>	-74	+07	-13	-10	-51	+01	+03	+10	-09
C	<u>+36</u>	<u>+28</u>	<u>+28</u>	+07	+20	+05	-03	+12	+07	+03	-06	+24	-17	+04	+04	-12	+06
D	+01	<u>+65</u>	<u>+65</u>	+00	+20	-11	+08	+02	-04	+36	-05	-12	+06	-10	+03	-49	+05
E	-27	+14	+14	+00	-03	-01	-07	-02	-18	-51	+01	-04	-04	-33	-03	-26	-10
F	+10	+10	+10	<u>+49</u>	-03	+15	+45	+04	-06	-07	-07	-14	-17	-13	-11	+00	+29
G	-16	-01	-01	-05	-03	+11	-31	+00	-04	<u>+58</u>	+02	+06	+03	-04	-17	-04	-01
H	-10	-06	-06	-09	<u>+36</u>	-12	+03	+06	+10	+01	+06	+12	-04	-10	+37	+58	-01
I	-17	-13	-13	-07	-02	-01	+72	+00	+00	+08	+12	+13	-11	+00	-07	+00	-07
J	-12	+10	+10	-05	+05	-07	-07	+01	+01	+05	<u>+66</u>	+02	-06	+05	-34	+03	-10
N	+19	-09	-09	-27	-17	-22	-26	<u>+47</u>	-25	-05	-07	-13	+12	-15	-02	-14	+02
O	+04	+04	+04	+03	+09	+03	-04	-16	+04	-08	+03	-01	-00	+08	+30	-04	+06
Q <sub>a</sub>	-29	+06	+06	-01	+09	-08	-12	+10	-03	+04	-05	+02	<u>+40</u>	+01	-12	+03	-33

	A	C	D	F	H	In <sub>L</sub>	In <sub>q</sub>	A <sub>q</sub>	G <sub>L</sub>	G <sub>q</sub>	J <sub>q</sub>	C <sub>L</sub>	O <sub>q</sub>	I <sub>3</sub>	O <sub>q</sub>	D <sub>q</sub>	I <sub>6</sub>
A																	
C																	
D																	
F																	
H																	
In <sub>L</sub>																	
In <sub>q</sub>																	
A <sub>q</sub>																	
G <sub>L</sub>																	
G <sub>q</sub>																	
J <sub>q</sub>																	
C <sub>L</sub>																	
O <sub>q</sub>																	
I <sub>3</sub>																	
O <sub>q</sub>																	
D <sub>q</sub>																	
I <sub>6</sub>																	

ages of 65, 73, 77, and 54 in the hyperplane.)

However, a hitherto undescribed phenomenon is encountered here, namely, the appearance of factors restricted to one medium, and appearing in one or both of the separate media alongside, and simultaneous with, the appearance of the joint medium factor having the same personality meaning. This is illustrated by C and C<sub>L</sub>, D and D<sub>q</sub> (Table 2), wherein the real psychological factor (C or D), loading the four essential variables across both media, carries alongside it an incomplete image of itself in each medium. The incomplete image loads only the two variables which belong in one medium. To these patterns, occurring simultaneously with the combined pattern, I have tentatively given the name "refraction factors," since they are analogous to what would be seen if one looked at an object both directly and refracted through a prism of another medium, one on each side of the line of vision.

Actually Table 2 does not simultaneously present *all* refraction factors for all real factors, but this should not disturb us any more than the failure of a single archeological digging to provide all the bones of a skeleton or all cultural elements for a given period. For, as it has been argued elsewhere (Cattell, 1958) any matrix typically has strictly as many dimensions as variables, and probably even more hyperplanes, i.e., one is always taking a selection in simple structure among more possible hyperplanes than one has chosen to extract factors. Further search should be made for refraction factors, therefore.

A vital empirical question affecting further inference at this point concerns the correlations among the real



and refraction factors for a given psychological dimension. We had expected them to be positively correlated, but the best estimate from existing data is that they are only slightly correlated, if at all. It is possible, however, that if more dimensions had been taken out their correlations would have been increased (see Diagram 5, Cattell, 1958).

Exploration and evaluation of possible hypotheses to account for refraction factors would require at least an article to itself. One does not go too far in interpretation, however, to say that they imply that each individual, in addition to his assessment on the real factor, gets a "bonus" on the variables peculiar to each medium, which is substantially unrelated to his status on the real factor. Our hypothesis is that these refraction factors belong to the perceptual class (Class 4 on page 166 above) and arise from the behavior in question being differently perceived in the two media. In self-rating a varying sensitivity and self-awareness—only in special cases a function of the trait being rated—could provide the differing "bonus" from person to person. The differing visibilities of these individuals from the position of the rater, giving the L-data refraction, would be expected to be quite unrelated to the order of their individual sensitivities in self-rating.

If this is correct one might also expect the lesser loadings, on variables other than the two salients, to be systematically different on the two refraction factors. For example, the rating by others, in the case of a factor much concerned in delinquency, might impart something of the stereotype of a scoundrel, where the Q-data refraction factor might convey more of a good person in difficulties. Since our main concern is

with the order which emerges little has been said of the "debris" factors notably 13 and 16 in Table 2. But our conclusion, tentatively, is that "evaluative" and "visibility" factors other than refraction factors are at present run together in the insufficient factor space so far used, and that, especially in the L data, these "halo" and related factors are substantial. They do not appear to be any known second-order factors, which can sometimes appear in inadequate first-order factorings. It has sufficed for our present investigation simply to set them aside. But if closer research scrutiny in this heap shows that our present indications are correct that these Class 4 perturbers are much larger in L than Q data, then the practice of trying to force questionnaire factors to align with rating "criteria" comes still more in question than it is today.

That the reader may more directly evaluate the nature and quality of the simple structure in Table 2 we have set out in Figure 2 a plot of two psychological ("real") factors therefrom.

#### SUMMARY AND CONCLUSION

1. Correlations among primary personality factors in different media do not provide a simple pattern of one-to-one relations, and fall decidedly short of unity between two factors of the same apparent psychological meaning.

2. The theoretical possibilities and the natural occurrences of perturbing influences hiding true alignment have been discussed and demonstrated. They have been classified as (a) test instrument factors; (b) actual trait modification by differing experience in subareas, requiring unity to be sought at a higher order level; (c) differences of density of representation of variables in different media;

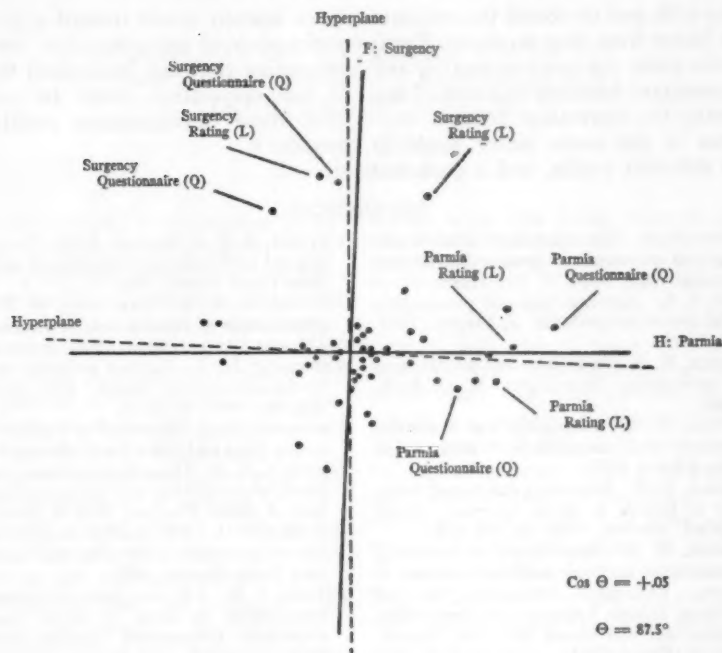


FIG. 2. Simple structure appearing between cross-media personality factors. (Marker variables labeled)

and (d) perceptual-evaluation-projection factors, occurring where human transmission of observations is involved.

3. Experimenters, especially when leaving rotation decisions to falsely founded analytical computer programs, commonly miss instrument factors, but when these are properly isolated and set aside by careful experiment it is possible to find the well known primary personality factors, each appearing as a single factor expressing itself in both L and Q media.

4. Regard for instrument and second-order-first-order factor relations is already producing clarity and consistency in personality structure research; but much remains to be explored regarding at least four forms of distortion which apparently occur

where human transmission is involved, i.e., in L and Q data. The new phenomenon of refraction factors particularly calls for intensive research.

5. One must distinguish between the question "Does a single simple structure factor exist loading variables of the same meaning on both media?" and "Can one get a perfect correlation between estimates of apparently (by meaning) the same factor, made in the two media?" Even when the answer to the first, so important for personality theory, is "Yes," as this paper claims to have shown, the answer to the second remains "No." The variance due to instrument factors, refraction factors, and any evaluation-perceptual factors peculiar to one medium will re-

main with and confound the estimate of a factor from that medium. Possibilities exist, by ipsative scoring and discriminant function methods of improving the correlation between estimates of the same factor made in two different media, and a path has

been opened above toward a proper estimation of the correction for attenuation that can be applied to see if the correlation could be unity. But these developments await research.

## REFERENCES

- BECKER, W. C. The matching of behavior rating and questionnaire personality factors. *Psychol. Bull.*, 1960, 57, 201-212.
- BERG, I. A. Response bias and personality: The deviation hypothesis. *J. Psychol.*, 1955, 40, 61-72.
- CATTELL, R. B. *Description and measurement of personality*. New York: World Book, 1946.
- CATTELL, R. B. *Personality and motivation structure and measurement*. New York: World Book, 1957.
- CATTELL, R. B. Extracting the correct number of factors in factor analysis. *Educ. psychol. Measmt.*, 1958, 18, 791-838.
- CATTELL, R. B. *Experiments on sources of perturbation in factor analytic resolution of traits*. (Advanced Publication No. 11) Urbana, Illinois: Laboratory of Personality Assessment and Group Behavior, University of Illinois, 1960.
- CATTELL, R. B., & ANDERSON, J. C. *The IPAT Music Preference Test of Personality*. Champaign, Illinois: Institute for Personality and Ability Testing, 1953.
- CATTELL, R. B., & COAN, R. W. Child personality structure as revealed in teachers' behavior ratings. *J. clin. Psychol.*, 1957, 13, 315-327.
- CATTELL, R. B., & COAN, R. W. Personality dimensions in the questionnaire responses of six and seven year olds. *Brit. J. educ. Psychol.*, 1958, 28, 232-242.
- CATTELL, R. B., & GRUEN, W. The primary personality factors in eleven year old children by objective tests. *J. Pers.*, 1955, 23, 460-478.
- CATTELL, R. B., RADCLIFFE, J., & SWENEY, A. The objective measurement of motivation structure in children. *J. clin. Psychol.*, 1960, 16, 227-232.
- CATTELL, R. B., & SAUNDERS, D. R. Interrelation and matching of personality factors from behavior rating, questionnaire and objective test data. *J. soc. Psychol.*, 1950, 31, 243-260.
- CATTELL, R. B., & SAUNDERS, D. R. Beiträge zur Faktoren-Analyse der Persönlichkeit. *Z. exp. angew. Psychol.*, 1955, 7, 319-343.
- CATTELL, R. B., & SCHEIER, I. H. *The meaning and measurement of neurosis and anxiety*. New York: Ronald, 1961.
- CATTELL, R. B., & WARBURTON, W. W. *A compendium of objective tests in personality*. Urbana: Univer. Illinois Press, in press.
- CRONBACH, L. J. Further evidence on response sets and test design. *Educ. psychol. Measmt.*, 1950, 10, 3-31.
- CRONBACH, L. J. *Essentials of psychological testing*. (2nd ed.) New York: Harper, 1960.
- DINGMAN, H. F. The relation between coefficients of correlation and difficulty factors. *Brit. J. statist. Psychol.*, 1958, 9, 13-18.
- EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.
- FRENCH, J. W. *The description of personality measurement in terms of rotated factors*. Princeton: Educational Testing Service, 1953.
- HOLZMANN, W. H., & BITTERMAN, M. E. A factorial study of adjustment to stress. *J. abnorm. soc. Psychol.*, 1956, 52, 179-185.
- MAYESKE, G. Some associations of musical preference dimensions of personality. Unpublished doctoral thesis, University of Illinois, 1961.
- MESSICK, S., & JACKSON, D. N. Acquiescence and the factorial interpretation of the MMPI. *Psychol. Bull.*, 1961, in press.
- ROBERTS, A. O. H. "Artifactor" analysis: Some theoretical background and practical demonstrations. *J. Nat. Inst. Personnel Res.*, Johannesburg, 1959, 7, 168-188.
- THURSTONE, L. L. *Primary mental abilities*. Chicago: Univer. Chicago Press, 1938.
- TOLLEFSON, D. Response to humor in relation to other measures of personality. Unpublished doctoral thesis, University of Illinois, 1961.
- WHERRY, R. J., & GAYLORD, R. H. Factor pattern of test items and test as a function of the correlation coefficient: Content, difficulty, and constant error factors. *Psychometrika*, 1944, 9, 237-244.

(Received July 20, 1960)

## COMMENTS ON CATTELL'S PAPER ON "PERTURBATIONS" IN PERSONALITY STRUCTURE RESEARCH

WESLEY C. BECKER

*University of Illinois*

Cattell's reply to my earlier paper (Becker, 1960) questioning the validity of published statements of a one-to-one matching between L-data and Q-data factors concedes the inaccuracy of those statements (see the first point in his summary). However, in the process of developing a defense for his basic *theoretical* position, Cattell has distorted the nature of my arguments to the point that a further brief clarification is needed.

Cattell states several times in his paper (Cattell, 1961) that since the evidence did not support his theory, I concluded that the evidence disproved his theory. In rebuttal I need only quote two sentences from my earlier paper.

It is apparent that the present evidence does not support the claim for "secure linkage" of BR and Q factors. This does not nec-

essarily imply that future research using more reliable and factor pure measures may not still prove Cattell's proposition to be correct (p. 208).

My critique was based on a question of fact, not of theory. Cattell has conceded this question of fact, as he must, but then he sets up for attack a question of theory which I did not raise. I did go on to indicate on *logical* grounds why I felt complete confirmation of his theory was exceedingly unlikely, and I see nothing in his present paper to change this opinion. The demonstration of a few "matchings" in the extraversion area, where on psychological grounds one would most expect self-perceptions and behavior ratings to overlap, can hardly be accepted as firm evidence for his general theoretical position.

### REFERENCES

- BECKER, W. C. The matching of behavior rating and questionnaire personality factors. *Psychol. Bull.*, 1960, **57**, 201-212.
- CATTELL, R. B. Theory of situational, instrument, second order, and refraction factors in personality structure research. *Psychol. Bull.*, 1961, **58**, 160-174.

(Received December 23, 1960)

## CATTELL REPLIES TO BECKER'S "COMMENTS"

RAYMOND B. CATTELL  
*University of Illinois*

In his additional comments, Becker expressly concedes that my theory has not been disproved. It is still odd that he objects to my saying that he considered the theory untrue, since he again says that it is "exceedingly unlikely," and, to a scientist, "true" and "untrue" mean "highly probable" or "highly improbable,"—at least, since the time of Victorian physics.

The positive conceptual and experimental contributions of my paper appearing since his comments, he either misses or ignores, since they show: (a) that it was impossible for him to reach any intelligible conclu-

sion on the theory without recognizing and developing the necessary corrections for attenuation and perturbation, and (b) that the facts which he says I must and do recognize are those chosen by Becker from experiments with older techniques. Science moves on, and the new facts which I present from technically more advanced designs show that *the same factor simultaneously loads on the hypothesized markers for both the rating and the questionnaire factors*. His statement that I concede his facts is therefore ambiguous.

(Received December 26, 1960)



---

*A complete library of psychology  
in one volume—*

# CLASSICS IN PSYCHOLOGY

Edited by Thorne Shipley

In a text of over 1400 pages, are the works of 35 leading psychologists of Europe, the United States, and the Soviet Union, spanning a period of one hundred and fifty years.

*Some of the chapters:*

<b>PHILIPPE PINEL:</b>	Treatise on Insanity
<b>JEAN MARTIN CHARCOT:</b>	Diseases of the Nervous System
<b>WILHELM WUNDT:</b>	Sensory Perception
<b>ERNEST MACH:</b>	The Relation of the Physical to the Psychical
<b>WILLIAM JAMES:</b>	Principles of Psychology
<b>EMIL KRAEPELIN:</b>	Clinical Psychiatry
<b>E. B. TITCHENER:</b>	Structural Psychology
<b>BENJAMIN RUSH:</b>	The Diseases of the Mind
<b>MANFRED SAKEL:</b>	Shock Treatment of Schizophrenia
<b>EUGEN BLEULER:</b>	Dementia Praecox
<b>ALFRED ADLER:</b>	Individual Psychology
<b>I. P. PAVLOV:</b>	Conditioned Reflexes
<b>HERMANN ROSSCHACH:</b>	Psychodiagnostics
<b>AUGUST AICHWORN:</b>	Wayward Youth
<b>G. STANLEY HALL:</b>	Adolescence
<b>JAN PIAGET:</b>	The Language and Thought of the Child
<b>KURT KOFFKA:</b>	Perception
<b>ISAAC RAY:</b>	Jurisprudence of Insanity
<b>WILLIAM McDUGALL:</b>	Social Psychology
<b>JOSEF BREUER &amp; SIGMUND FREUD:</b>	Studies on Hysteria
<b>C. G. JUNG:</b>	Analytical Psychology

\$20.00

*You can expedite shipment by enclosing remittance.*

**PHILOSOPHICAL LIBRARY, Publishers**

15 East 40th Street, New York 16, N.Y.

---



# Elements of Psychology



by DAVID KRECH and RICHARD S. CRUTCHFIELD,  
*University of California*

"In its coverage of the major topics of psychology, there is depth as well as breadth. All material is lucidly organized and intelligently integrated. Sketches and diagrams are used freely, skillfully, and tastefully to illustrate empirical findings, details of experimental apparatus, and neurological structure relevant to perception and emotion."—*Psychiatric Quarterly*, 753 pp. illus. \$7.00.

THE STUDENT WORKBOOK by James L. McGaugh, priced at \$2.00, is available supplement to *Elements of Psychology*. An alternate edition of the INSTRUCTOR'S MANUAL is now available.

## Clinical and Counseling Psychology

by JOHN M. HADLEY, *Purdue University*

"An excellent practical book which organizes the experience of the reader into meaningful wholes and for once gives the student an inclusive introduction to himself as a counselor or clinician."—GRAHAM B. HELL, *Pomona College*. "The book is relatively comprehensive and mature. Topics have been judiciously selected, intelligently ordered and amply discussed; and the book's organization is particularly interesting."—*Psychiatric Quarterly*, 720 pp. \$7.00.

ALFRED A. KNOPF, *Publisher*

College Department

501 Madison Avenue

New York 22

GEORGE BANTA COMPANY, INC., MENAHEA, WISCONSIN, U.S.A.

